Concise and tight security analysis of the Bennett-Brassard 1984 protocol with finite key lengths

Masahito Hayashi[1,2] **and Toyohiro Tsurumaru**[3]

[1] Graduate School of Mathematics, Nagoya University, Furocho, Chikusa-ku, Nagoya, 464-860 Japan

[2] Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117542

[3] Mitsubishi Electric Corporation, Information Technology R&D Center, 5-1-1 Ofuna, Kamakura-shi, Kanagawa, 247-8501, Japan

Abstract. We present a tight security analysis of the Bennett-Brassard 1984 protocol taking into account the finite size effect of key distillation, and achieving unconditional security. We begin by presenting a concise analysis utilizing the normal approximation of the hypergeometric function. Then next we show that a similarly tight bound can also be obtained by a rigorous argument without relying on any approximation.

In particular, for the convenience of experimentalists who wish to evaluate the security of their QKD systems, we also give explicit procedures of our key distillation, and also show how to calculate the secret key rate and the security parameter from a given set of experimental parameters. Besides the exact values of key rates and security parameters, we also present how to obtain their rough estimates using the normal approximation.

1. Introduction

The finite size effect is an important issue in practical quantum key distribution (QKD) systems. The first detailed finite-size analysis for general coherent attacks was given by Hayashi [1] using the normal approximation. Later, Scarani and Renner [2] gave a simple analysis based on the quantum de Finetti Theorem, but their results are valid only against collective attacks. Matsumoto and Uyematsu also gave a simple analysis [3], but again, essentially valid only for collective attacks. Later, Tomamichel et al. [4] gave a tighter bound with unconditional security by using the uncertainty relations (see., e.g., [5, 6]).

In this paper, we present a concise analysis for the Bennett-Brassard 1984 (BB84) protocol [7] that takes the finite key effect into account and yields better key generation rates, with and without relying on the normal approximation. Our analysis is valid for general coherent attacks and thus our results guarantee the unconditional security. For the sake of simplicity, we consider the case where the sender, Alice, has a perfect single photon source. We also assume that Alice and the receiver, Bob, calculate an upper bound on the phase error rate of a sifted key, from that of the corresponding sample bits; hence the key generation rate can vary each time Alice and Bob run of the protocol.

Throughout the paper we use the security criteria with universal composability; the same criteria as used by many researcher, particularly by Renner and his coworkers [8, 9]. Hence our final goal is to show that the trace distance between the actual and the ideal states can be bounded from above. However, in the mathematical analysis for obtaining upper bounds on the trace distance, we do not use Renner's approach based on the smooth minimum entropy [8]. Instead, we bound the trace distance using the argument by Shor and Preskill [10], as well as its modification by Hayashi [1]. In Section 3, by using these formalisms, we show that the trace distance can be bounded by using the decoding error probability $P_{\rm ph}$ of the virtual phase error correction; in other words, the universally composable security can be guaranteed by bounding $P_{\rm ph}$. To the best of our knowledge, our argument here is the first rigorous treatment of the universally composable security based on the Shor-Preskill formalism, applicable to linear universal hash functions with variable final key lengths.

As we shall also discuss at the end of Section 3, in order to achieve high key generation rates and strong bounds on $P_{\rm ph}$ simultaneously, it is crucial to estimate the phase error rate $p_{\rm sft}$ of the sifted key with a high accuracy. Note here that the quantity $p_{\rm sft}$ cannot be measured directly in the BB84 protocol. Hence in Section 4, we solve an interval estimation problem on $p_{\rm sft}$ using the hypergeometric distribution $P_{\rm hg}$. Then by using the obtained result, we give explicit bounds on $P_{\rm ph}$ in Section 5. In particular, in order to clarify the argument, we present two versions of analysis: We first derive a simple bound that we call the *straightforward bounds* (Propositions 1 and 2); and then next give a more complicated bound called the *Gaussian bounds* (Theorems 2 and 3), which yield a better final key rate if the raw key is sufficiently large. For the both types of bounds, we first present a simple analysis based on the normal approximation of the

hypergeometric function (Proposition 1 and Theorem 2), and then next show that a similarly tight bound can also be obtained by a rigorous argument without relying on any approximation (Proposition 2 and Theorem 3).

Since this paper is not aimed only at theorists, but also at experimentalists who wish to evaluate the security of their QKD systems, we include explicit procedures of security evaluation. We begin in Section 2 by explaining explicit procedures of our key distillation. Then after theoretical arguments of the security, we demonstrate in Section 6 how to use our theorems to calculate the secret key rate and the security parameter (i.e., an upper bound on the trace distance) from a given set of experimental parameters. Besides the exact values of key rates and security parameters, we also present how to obtain their rough estimates using the normal approximation.

In order to show that our rates are indeed better than in existing literatures, e.g., Refs. [2, 4], we draw in Section 7 example curves of key generation rates (Figs. 1 and 2). There are several reasons for this improvement. First, our upper bounds are close to the approximated value of the hypergeometric distribution obtained by the normal approximation, while the existing results [2, 4] did not discuss the closeness to the normal approximation. Second, in our method, the adversary's information is estimated in terms of the Shannon entropy, whereas in [2, 4] they use the minimum entropy, which is a lower bound on the Shannon entropy. Finally, we use an error margin that depends on the measured error rates of sample bits, while in Refs. [2, 4] the margin is a constant.

We also treat the sacrifice bit length with the second order coding rate, which draws the attention from information theory community [11, 12, 13]. The conventional asymptotic theory treats the coding length with the first order coefficient. It is impossible to treat the approximation value of the best error probability with the first order coefficient of the coding length. However, it becomes possible if we consider the coding length up to the second order coefficient. In this paper, we derive an asymptotic approximation value of the upper bound of the universally composable security criterion when the sacrifice bit length is given as the form $nh(p_{\rm smp})+\sqrt{n}g(p_{\rm smp})$ with the measured phase error rate, where a function $g(p_{\rm smp})$ of $p_{\rm smp}$ will be given with a concrete form in Section 4 (Theorem 4).

The differences from our previous papers are as follows. In Refs. [1], Hayashi simply approximated the hypergeometric distribution by the normal distribution having the same variance, without showing its validity. In this paper, we present a rigorous analysis without relying on any approximation (Proposition 2 and Theorem 3), by using upper bounds on the hypergeometric distribution obtained from the Stirling's formula and inequalities proved in Ref. [14, 15]. As mentioned above, we also included the first rigorous treatment of the universally composable security based on the Shor-Preskill formalism, applicable to linear universal hash functions with variable final key lengths.

2. Description of Our QKD Protocol

We consider the following type of the BB84 protocol. This protocol differs from existing versions (e.g., [1, 2, 3]) only in the phase estimation and the privacy amplification steps.

Generation of a Sifted Key and Sample Bits Alice and Bob start the protocol with a quantum communication and obtain a sifted key of n bits and sample bits of l bits. Here we assume that raw key bits are chosen from the uniform distribution. The sample bits must be selected randomly, and a sifted key and the sample bits must be measured in different bases.

For example, suppose that Alice and Bob exchange N qubits, choosing the x basis with probability q, and the z basis with $1-q$. Then, on average, Nq^2 bits coincide in the x basis, and $N(1-q)^2$ in the z basis. By assinging the x basis for a sifted key, and the z basis for sample bits, they have $n = Nq^2$, $l = N(1-q)^2$.‡

Bit Error Correction Bob corrects bit errors in his sifted key using a linear error correcting code. For example, as in Shor-Preskill's case [10], Alice may announce a random bit string XORed with her sifted key; or alternatively, as in Koashi's case [16], she may send a syndrome of her sifted key encrypted with a previously shared secret key. In either case, Alice and Bob end up with $n(1-fh(p_{\mathrm{bit}}))$ bits of reconciled key k_{rec}, with the bit error rate p_{bit} of a sifted key. Here $h(x)$ is the binary entropy function defined as $h(x) := -x\log_2 x - (1-x)\log_2(1-x)$, and value f corresponds to the efficiency of the error correcting code used. For practical codes, $f \simeq 1.1$. It should be noted that here the sizes of bit error correcting codes are independent of the security, and thus Alice and Bob may perform bit error correction by dividing a sifted key k_{sif} of n bits to arbitrarily smaller blocks.

In many cases, one needs to guarantee the correctness of the shared keys, that is, one has to minimize the probability ϵ_{cor} that Alice's and Bob's secret keys do not match and the protocol does not abort. One way for minimizing ϵ_{cor} is that Alice calculates an r-bit hash value of her reconciled key k_{rec} using universal$_2$ hash functions. Then she encrypts it with the one-time pad using a previously shared secret key, and sends it to Bob. Bob also calculates his own hash value, and if it does not match Alice's, they abort the protocol§. By doing this, we have $\epsilon_{\mathrm{cor}} \leq 2^{-r}$.

Estimation of the number of phase errors in the channel In order to use privacy amplification properly and guarantee the security of a secret key, Alice and Bob need to know an upper bound on the number of phase errors occurring in the channel. It

‡ In general, however, Alice and Bob may choose bases with different probabilities, and a sifted key and sample bits may be chosen with arbitrary proportions from the two basis.

§ Another possibility is to continue protocol by exchanging supplementary information, such as an additional syndrome, over the public channel, and try bit error correction again. In such case, the supplementary information also needs to be encrypted with a formerly shared key.

should be noted here that the phase error is a completely different concept from the bit error mentioned above (for details, see Section 3). Since the phase error rate cannot be measured directly in practical QKD systems, we estimate its upper bound from the measured error rate of samples.

We denote the number of bit errors occurring in a sample bits by c, and the corresponding bit error rate by $p_{\mathrm{smp}}(c) := c/l$. We also call the union of a sifted key and the sample bits *total bits*, and denote the number of their bit errors by k. Hence the error rate of total bits is given by $p(k) := k/(n+l)$, and that of a sifted key by $p_{\mathrm{sft}}(k,c) = (k-c)/n$. Note here that measuring c corresponds to randomly sampling phase errors in the total bits, because a sifted key and the samples are measured in different bases. Due to this fact, the measured value of $p_{\mathrm{smp}}(c)$ is used to estimate an upper bound on $p_{\mathrm{sft}}(k,c)$. In the asymptotic limit $n, l \to \infty$, Alice and Bob may assume $p_{\mathrm{sft}}(k,c) = p_{\mathrm{smp}}(c)$. In practical QKD systems, however, the two values differ in general due to statistical fluctuations. Thus they obtain a statistically estimated upper bound of $p_{\mathrm{sft}}(k,c)$ as a function of the measured value c, which we denote by $\hat{p}_{\mathrm{sft}}(c)$. Throughout the paper, we make it a rule to denote an estimated upper bound of a random variable v by $\hat{v}$. The explicit functional form of $\hat{p}_{\mathrm{sft},\varepsilon}(c)$ is discussed later, and is given in Eq. (25).

Privacy Amplification (PA) The estimated phase error rate $\hat{p}_{\mathrm{sft}}(c)$ can be used to obtain an upper bound the amount of information that is leaked to Eve. In order to cancel Eve's information, Alice and Bob perform a classical data processing called privacy amplification on the reconciled key k_{rec} to generate the secret key k_{sec}; very roughly speaking, PA randomizes and shrinks k_{rec} so that Eve's information is canceled by the remaining fraction that is unknown to Eve. The number of bits to be reduced in this process (sacrifice bits) is determined from $\hat{p}_{\mathrm{sft}}(c)$ in the following manner.

We set two limits $c_{\min}$ and $c_{\max}$ ($c_{\min} \le c_{\max}$) on the sample bit error c, depending on which Alice and Bob change their procedures.

- If $c_{\max} < c$, Alice and Bob abort the protocol.
- If $c_{\min} \le c \le c_{\max}$, Alice and Bob generate a secret key as the hash value of their sifted key by using a linear and surjective universal$_2$ hash functions. The number $\alpha(c)$ of sacrifice bits, i.e., the number of bits reduced in PA, is given by

$$\alpha(c) = n\lceil h\left(\hat{p}_{\mathrm{sft},\varepsilon}(c+2)\right)\rceil + D.$$

  Here $\lceil x \rceil$ denotes the smallest integer larger than or equal to x. Hence, as a result, they obtain a secret key k_{sec} of $G = n\left[1 - fh(p_{\mathrm{bit}})\right] - \lceil nh\left(\hat{p}_{\mathrm{sft},\varepsilon}(c+2)\right)\rceil - D$ bits. ‖

‖ Note that key length G of (2) differs from the asymptotic case ($l, n \to \infty$) essentially only in the definition of phase error rate $\hat{p}_{\mathrm{sft},\varepsilon}(c+2)$. Hence the estimation of $\hat{p}_{\mathrm{sft},\varepsilon}(c+2)$ is the key point of our finite size analysis.

- If $c < c_{\min}$, Alice and Bob generate a secret key in the same way as above, except that they sacrifice $\alpha(c) = \lceil nh\left(\hat{p}_{\mathrm{sft},\varepsilon}(c_{\min}+2)\right)\rceil + D$ bits for PA. As a result, they obtain a secret key k_{sec} of $G = n\left[1 - fh(p_{\mathrm{bit}})\right] - \lceil nh\left(\hat{p}_{\mathrm{sft},\varepsilon}(c_{\min}+2)\right)\rceil - D$ bits.

Alternatively, we can combine these three case as follows: Define the sacrificed bit length $\alpha(c)$ to be

$$\alpha(c) = \lceil nh\left(\hat{p}_{\mathrm{sft},\varepsilon}\left(\max[c, c_{\min}]+2\right)\right)\rceil + D. \tag{1}$$

If $c \leq c_{\max}$, Alice and Bob sacrifice $\alpha(c)$ bits for PA and obtain the final key of length

$$G(c) = n\left[1 - fh(p_{\mathrm{bit}})\right] - \alpha(c). \tag{2}$$

If $c \geq c_{\max}$, they abort the protocol.

In practice, the most efficient implementation of PA is to use the Toeplitz matrices: Alice and Bob select a bit-valued Toeplitz matrix M randomly by communicating over the public channel, multiply it with a reconciled key k_{rec} modulo 2, and obtain the secret key $k_{\mathrm{sec}} = Mk_{\mathrm{rec}}$ (for details, see., e.g., [8, 17, 18]).

In this paper, we additionally require the surjectivity for all of hash functions. To the best of our knowledge, the most efficient implementation of linear and surjective universal$_2$ functions is by using the modified Toeplitz matrix, introduced in [1, 17]; in this case we replace M above by a concatenation $M' = (I, T)$ of the (square) identity matrix I and a Toeplitz matrix T. Note that this modification M' is slightly more efficient than M above. Also note that unlike M', the normal Toeplitz matrix M gives a non-surjective map with a very small but nonzero probability; e.g., for M being an all-zero or all-one matrix.

It should be noted here that, unlike in bit error correction, one is not allowed to perform PA by dividing k_{rec} and k_{sec} into smaller blocks, because doing so will destroy the universal$_2$ property of the (modified) Toeplitz matrix. Also note here that the both key lengths, $|k_{\mathrm{rec}}| = n[1 - fh(p_{\mathrm{bit}})]$ and $|k_{\mathrm{sec}}| = G$, are of order $O(n)$. If one applies a naive multiplication algorithm, the computational complexity (i.e., the processing time) increases as $O(n^2)$ (i.e., $O(n)$ per key), and thus becomes a severe bottle neck of the key distillation. This is in fact the most explicit impact of the finite size effect on practical QKD systems.

One way around this problem is to use an efficient multiplication algorithm for a Toeplitz matrix and a vector exploiting the fast Fourier transform (FFT) algorithm (see, e.g., [19]). The complexity of this efficient algorithm scales as $O(n \log n)$, or $O(\log n)$ per bit, which can be regarded as a constant in practice. An actual implementation shows that the throughput exceeds 1Mbps for $|k_{\mathrm{rec}}| = 10^6$ on software, as demonstrated, e.g., in Ref. [18].

	total bits	sifted key	sample bits
Number of bits	$n+l$	n	l
Number of errors	k	$k-c$	c
Error rate	$p(k)=\frac{k}{n+l}$	$p_{\rm sft}(k,c)=\frac{k-c}{n}$	$p_{\rm smp}(c)=\frac{c}{l}$
Estimate of error rate with error probability ε	$\hat{p}_\varepsilon(c)$	$\hat{p}_{{\rm sft},\varepsilon}(c)$	

Table 1. Notations of the key lengths, total bits, and sample bits. Functions $\hat{p}_\varepsilon(c)$ and $\hat{p}_{{\rm sft},\varepsilon}(c)$ denote the estimated upper bounds of $p(k)$ and $p_{\rm sft}(k,c)$, under the condition that there are c errors in sample bits. Parameter ε denotes the probability that the estimation fails. See Section 4 for details.

3. Security Criteria of the BB84 Protocol in the finite case

3.1. The security of QKD with universal composability

We employ the definition of the security of QKD with universal composability in the variable length case [20]. In order to guarantee the security for our protocol, we need to evaluate the security criteria with universal composability after the privacy amplification [9]. In this paper, we apply the above definition with the variable length case to the final state after the privacy amplification [21].

For this purpose, we describe all public information by x, including the choice of a hash function (which corresponds, e.g., to "f" of [9]), and the length of the final key (e.g., "m" of [20]). However, here we do not restrict ourselves with those two cases; it may contain other public information, e.g., the choice of a code for bit error correction. Hence the length m of the final key is of course a function of x. We denote the probabilistic distribution of x in the actual protocol by $P_{\rm pub}(x)$.

Then we consider the Hilbert space $\mathcal{H}_A\otimes\mathcal{H}_E\otimes\mathcal{H}_X$, consisting of Alice's final key $\mathcal{H}_A$, Eve's system $\mathcal{H}_E$, and the public information $\mathcal{H}_X$. We define $\mathcal{H}_A=(\mathbb{C}^2)^M$ with M sufficiently large; so that when $m(x)<M$, Alice uses the (preassigned) subspace of $\mathcal{H}_A$. Also, following [8], we define the composite system of E and X to be E', i.e., $\mathcal{H}_{E'}=\mathcal{H}_E\otimes\mathcal{H}_X$. We denote by $\rho_{A,E|x}$ the state of Alice and Eve after privacy amplification, conditioned on public information x. Hence, the state after privacy amplification takes the form $\rho_{A,E'}=\sum_x P_{\rm pub}(x)\rho_{A,E|x}\otimes|x\rangle\langle x|$.

In this notation, we consider conditional probabilities with respect to length m of the final key. The actual protocol generates the final key of m bits with probability $P_{\rm len}(m):=\sum_{x:m(x)=m}P_{\rm pub}(x)$. The public information x obeys the conditional distribution $P(x|m):=\frac{P_{\rm pub}(x)}{P_{\rm len}(m)}$; hence the conditional actual state given m is a density matrix $\rho_{A,E'|m}:=\sum_{x:m(x)=m}P_{\rm pub}(x|m)\rho_{A,E|x}\otimes|x\rangle\langle x|$. The corresponding ideal state given m is defined to be $\rho_{{\rm Ideal}\,|m}:=\rho^{\rm mix}_{A|m}\otimes\rho_{E'|m}$, where $\rho^{\rm mix}_{A|m}$ is the completely mixed state in the m-qubit subsystem of $\mathcal{H}_A$, and $\rho_{E'|m}:={\rm Tr}_A\rho_{A,E'|m}$. Thus, under the condition that the final key length is m, the universal compsable security can be guaranteed by bounding the trace distance of these two states, i.e., $\left\|\rho_{A,E'|m}-\rho_{{\rm Ideal}\,|m}\right\|_1$ [9].

Parameter m is a random variable in our protocol; hence following [20], we define the universally composable security by bounding the average trace distance $\sum_m P_{\rm len}(m) \left\| \rho_{A,E'|m} - \rho_{\rm Ideal\,|m} \right\|_1$. In this case, it is convenient to define $\rho_{\rm Ideal} := \sum_m P_{\rm len}(m) \rho_{\rm Ideal\,|m}$. Then the average trace distance can be rewritten as

$$\begin{aligned} \left\| \rho_{A,E'} - \rho_{\rm Ideal} \right\|_1 &= \sum_m P_{\rm len}(m) \left\| \rho_{A,E'|m} - \rho^{\rm mix}_{A|m} \otimes \rho_{E'|m} \right\|_1 \\ &= \sum_x P_{\rm pub}(x) \left\| \rho_{A,E|x} - \rho^{\rm mix}_{A|m(x)} \otimes \rho_{E|x} \right\|_1 \qquad (3) \\ &\leq \sum_x P_{\rm pub}(x) \left\| \rho_{A,E|x} - \rho_{A|x} \otimes \rho_{E|x} \right\|_1 \qquad (4) \\ &\quad + \sum_x P_{\rm pub}(x) \left\| \rho_{A|x} - \rho^{\rm mix}_{A|m(x)} \right\|_1 , \qquad (5) \end{aligned}$$

where $\rho_{A|x} := {\rm Tr}_E \rho_{A,E|x}$. Hence one may instead bound the sum of the second and the third lines. Here we used the fact that $\rho_{A,E'} = \sum_x P_{\rm pub}(x) \rho_{A,E|x} \otimes |x\rangle\langle x| = \sum_m P_{\rm len}(m) \rho_{A,E'|m}$ for the first equality; and $\rho_{E'|m} = \sum_{x:m(x)=m} P_{\rm pub}(x|m) \rho_{E|x} \otimes |x\rangle\langle x|$ for the second equality. The quantity of (5) measures the non-uniformity of Alice's final key; i.e., it gives the averaged distance between Alice's partial state $\rho_{A|x}$ and the ideally mixed state $\rho^{\rm mix}_{A|m(x)}$. Note that these two states equal when Alice and Bob choose a surjective hash function, because we assume that Alice's raw key obeys the uniform distribution. In particular, if Alice and Bob use a hash function family which consists only of surjective functions (such as the modified Toeplitz matirices [1, 17] mentioned in the previous section), it suffices to bound (4) only.

3.2. Decoding error probability of the virtual phase error correction

We believe that the above definition of security based on the trace distance is the same as the one used by Renner and others [8, 9]. Throughout the paper we employ this definition of security. However, in the remaining part where we actually obtain upper bounds on the trace distance, we do not use Renner's approach based on the smooth minimum entropy [8]. Instead, we bound the trace distance $\| \rho_{A,E|x} - \rho_{A|x} \otimes \rho_{E|x} \|_1$ appearing in (4) using the well-known argument by Shor and Preskill [10], as well as its modification by Hayashi [1]. As we shall see shortly, in these formalisms, the trace distance is bounded from above by using the decoding error probability of the (virtual) phase error correction¶, which can be identified with the privacy amplification in the actual protocol. The first step of the proof is to consider a virtual protocol where Alice and Bob correct bit errors as well as phase errors occurring in the quantum channel (under Eve's influence) by using the Calderbank-Shor-Steane (CSS) code. By correcting these two types of errors, Alice and Bob can guarantee that their virtual channel (obtained as a result of quantum error correction) is noiseless and decoupled from Eve; thus the key they exchange there is unconditionally secure. The second step of the proof is to note that, from Eve's view point, this virtual protocol is completely

¶ The probability that the (virtual) decoding algorithm fails to give a correct answer.

indistinguishable from the actual protocol. By using this indistinguishability, the security of the actual protocol follows automatically from that of the virtual protocol.

In these formalisms, phase error correction in the virtual protocol is transformed to a simple classical data processing in the actual protocol. That is, Alice and Bob do not need to perform phase error correction in the actual protocol; instead it suffices to perform a projection $C_1 \to C_1/C_2$, where C_1, C_2 are the classical CSS code. The projection $C_1 \to C_1/C_2$ is often called privacy amplification (PA). This is why we often identify PA with the virtual phase error correction in this paper$^+$. (In Ref. [17], we have shown that the projection $C_1 \to C_1/C_2$ can be replaced by an ε-almost dual universal$_2$ hash function family.)

The original argument of Shor and Preskill was later improved in Refs. [22, 23], where it was shown that the virtual phase error correction and the bit error correction can be discussed separately. In fact the virtual phase error correction is essential for guaranteeing security, while the bit error correction is necessary only for equalizing Alice's and Bob's final keys. As a result of this observation, the trace distance $\|\rho_{A,E|x} - \rho_{A|x} \otimes \rho_{E|x}\|_1$ of (4) can be bounded as [1]

$$\left\|\rho_{A,E|x} - \rho_{A|x} \otimes \rho_{E|x}\right\|_1 \leq 2\sqrt{2}\sqrt{P_{\mathrm{ph}|x}}, \tag{6}$$

where $P_{\mathrm{ph}|x}$ denotes the conditional decoding error probability of the virtual phase error correction, given public information x. By taking the average of (6) with respect to x, and by noting that the function $a \mapsto \sqrt{a}$ is concave, we have

$$\sum_x P_{\mathrm{pub}}(x) 2\sqrt{2}\sqrt{P_{\mathrm{ph}|x}} \leq 2\sqrt{2}\sqrt{\sum_x P(x)_{\mathrm{pub}} P_{\mathrm{ph}|x}} = 2\sqrt{2}\sqrt{P_{\mathrm{ph}}}, \tag{7}$$

where P_{ph} denotes the decoding error probability of the virtual phase error correction.

As to the non-uniformity of the final key given in (5), recall that we assumed that Alice's random variable obeys the uniform distribution. Then the left over hash lemma [24, 25] yields

$$\sum_x P_{\mathrm{pub}}(x) \|\rho_{A|x} - \rho^{\mathrm{mix}}_{A|m(x)}\|_1 \leq \sum_x P_{\mathrm{pub}}(x) 2^{-\frac{\alpha(x)}{2}}, \tag{8}$$

where $\alpha(x)$ is the number of sacrifice bits in the privacy amplification.

Hence by combining (3)$\sim$(5), (7), and (8) we obtain

$$\|\rho_{A,E'} - \rho_{\mathrm{Ideal}}\|_1 \leq 2\sqrt{2}\sqrt{P_{\mathrm{ph}}} + \sum_x P_{\mathrm{pub}}(x) 2^{-\frac{\alpha(x)}{2}}. \tag{9}$$

In other words, in order to guarantee the security with universal composability, it suffices to bound the quantity on the right hand side of (9). In particular, as we have noted below (5), the second term on the right hand side of (9) is exactly zero when all of the hash functions are surjective; in this case the above inequality is replaced by

$$\|\rho_{A,E'} - \rho_{\mathrm{Ideal}}\|_1 \leq 2\sqrt{2}\sqrt{P_{\mathrm{ph}}}. \tag{10}$$

$^+$ However, the actual protocol does not necessarily have a counterpart for any operation in the virtual protocol. For example, the actual protocol has no operation corresponding to measurement of the syndrome in the virtual protocol.

Hence, in order to guarantee the universally composable security, it suffices to bound $P_{\rm ph}$.

3.3. Conditional decoding error probability given k

In this subsection we show that, in order to bound the decoding error probability $P_{\rm ph}$ of the virtual phase error correction, it is sufficient to bound $P_{{\rm ph}|k}$ for all k, where $P_{{\rm ph}|k}$ denotes the corresponding conditional probability given k. We also show that a bound on $P_{{\rm ph}|k}$ can be given in a concise form using the hypergeometric distribution $P_{\rm hg}(c|k)$ and binary entropies.

First note that, without loss of generality, Eve's eavesdropping strategy can be described by the probability distribution $Q_{\rm Eve}(k)$ of k, which is the number of errors in the total bits $n+l$*. Then $P_{\rm ph}$ can be rewritten as $P_{\rm ph} = \sum_k Q_{\rm Eve}(k) P_{{\rm ph}|k}$, where $P_{{\rm ph}|k}$ denotes the conditional decoding error probability given k.

Next we consider the conditional probability $P_{\rm hg}(c|k)$ of c given k; i.e., the probability that c bits of errors are found in sample bits when there are k errors in the total bits. Since sample bits are sampled without replacement, c obeys the hypergeometric distribution for a fixed value of k:

$$P_{\rm hg}(c|k) := \frac{\binom{n}{k-c}\binom{l}{c}}{\binom{n+l}{k}}, \tag{11}$$

with the average $\bar{c}$ and the deviation σ given by

$$\bar{c}(k) := \frac{lk}{n+l}, \quad \sigma_{n,l}(k)^2 := \frac{knl(n+l-k)}{(n+l)^2(n+l-1)}. \tag{12}$$

In the following, $\sigma_{n,l}(k)^2$ is simplified to $\sigma(k)^2$. Hence values of k, c occurs with probability $Q_{\rm Eve}(k)P_{\rm hg}(c|k)$. (Here sample bits are sampled without replacement simply because one cannot measure both the phase and the bit values of a qubit simultaneously, and thus Alice and Bob cannot reuse the sample bits as a sifted key. If one could somehow sample them with replacements, the hypergeometric distribution here would of course be replaced by the binomial distribution, which is much simpler.)

Finally we consider the conditional decoding error probability $P_{{\rm ph}|k,c}$ for fixed values of k and c. In this case, the number of phase error patterns of total bits is bounded from above by $2^{nh((k-c)/n)}$ (see, e.g., Lemma 4.2.2, Ref. [29]). Due to the construction of the procotocl, the number of the sacrificed bits $\alpha(c)$ is fixed. As we have shown in Ref. [17], if Alice and Bob use a linear universal$_2$ hash function family for PA in the actual protocol, it can be considered as the situation in the virtual protocol where they use a 2-almost universal$_2$ linear code family for phase error correction (i.e., a linear 2-almost universal$_2$ hash function family is used as the syndrome function for correcting phase errors). Then the decoding error probability $P_{{\rm ph}|k,c}$ of the virtual phase error correction

* In the general setting, Eve is allowed to use the superposition among different intgers k. In order to treat such a case, we introduce the distribution $Q_{\rm Eve}(k)$ here.

can be bounded as

$$\begin{aligned} P_{\mathrm{ph}|k,c} &\leq S_{\mathrm{pa}}(k,c) := 2 \cdot 2^{[g(k,c)]^-} = 2^{[g(k,c)]^- + 1}, && (13) \\ g(k,c) : &= nh\left((k-c)/n\right) - \alpha(c) \\ &= nh\left((k-c)/n\right) - nh\left(\hat{p}_{\mathrm{sft}}(c+2)\right) - D && (14) \\ &= nh\left(p_{\mathrm{sfc}}(k,c)\right) - nh\left(\hat{p}_{\mathrm{sft}}(c+2)\right) - D, \end{aligned}$$

where $[x]^- := \min(x, 0)$. It is easy to see that Inequality (13) holds when the completely random matrices (a type of universal$_2$ hash functions) are used for PA, as in Koashi's case [16]. It is also shown to hold when the Toeplitz matrices (another universal$_2$ hash function family) are used for PA, by using the fact that dual matrices of the Toeplitz matrices generate universal$_2$ hash functions [1]. More generally, in Ref. [17], we have further shown that Inequality (13) is valid when an arbitrary family of universal$_2$ functions is used for PA.

Hence, to summarize, under Eve's strategy $Q_{\mathrm{Eve}}(k)$, error numbers k, c are distributed by $Q_{\mathrm{Eve}}(k)P_{\mathrm{hg}}(c|k)$. For fixed values of k, c, the virtual phase error correction fails with a probability less than $S_{\mathrm{pa}}(k,c)$ given in (13). Combining these probabilities, we see that the decoding error probability P_{ph} of the virtual phase correction can be bounded as

$$\begin{aligned} P_{\mathrm{ph}} = \sum_k Q_{\mathrm{Eve}}(k) P_{\mathrm{ph|k}} &\leq \sum_k \sum_c Q_{\mathrm{Eve}}(k) P_{\mathrm{hg}}(c|k) S_{\mathrm{pa}}(k,c) && (15) \\ &= \sum_k Q_{\mathrm{Eve}}(k) S_{\mathrm{av}}(k) \leq \max_k S_{\mathrm{av}}(k), && (16) \end{aligned}$$

where $S_{\mathrm{av}}(k)$ is defined by

$$S_{\mathrm{av}}(k) := \sum_{c=0}^{c_{\max}} P_{\mathrm{hg}}(c|k) S_{\mathrm{pa}}(k,c). \tag{17}$$

Since Eve's strategy $Q_{\mathrm{Eve}}(k)$ can be arbitrary, P_{ph} can be bounded if and only if $\max_k S_{\mathrm{av}}(k)$ is bounded. Hence in what follows, we will concentrate on obtaining upper bounds on $\max_k S_{\mathrm{av}}(k)$.

As one can see from the definition of $S_{\mathrm{pa}}(k,c)$ in (13), (14), a straightforward way of minimizing $\max_k S_{\mathrm{av}}(k)$ is to define the function $\hat{p}_{\mathrm{sft}}(c)$ so that it always gives a large value; this corresponds to the situation where, looking at c, Alice and Bob always give a pessimistic estimate $\hat{p}_{\mathrm{sft}}(c)$ that is much larger than the actual value $p_{\mathrm{sft}}(k,c)$. However, as one can see from the definition of $\alpha(c)$ in (1) and the final key length G given in the previous section, a large $\hat{p}_{\mathrm{sft}}(c)$ results in a poor key generation rate. Rather, in order to achieve high key generation rates and the high-level security simultaneously, one needs to minimize $\max_k S_{\mathrm{av}}(k)$ by considering the contributions of the two factors, $P_{\mathrm{hg}}(k|c)$ and $S_{\mathrm{pa}}(k,c)$. Hence we define $\hat{p}_{\mathrm{sft}}(c)$ so that it becomes as close as possible (and larger) to the actual value $p_{\mathrm{sft}}(k,c)$, in the regions of k, c where $P_{\mathrm{hg}}(c|k)$ is not negligible. This is equivalent to the estimation problem of an upper bound of $p_{\mathrm{sft}}(k,c)$:

(i) For a given c, we give a suitable choice of the estimated value $\hat{p}_{\rm sft}(c)$ for the phase error rate of a sifted key. Alice and Bob use this value to calculate the value of $\alpha(c)$ of (1), and obtain the final key length G. This will be done in Section 4.

(ii) With the suitable choice of $\hat{p}_{\rm sft}(c)$, we obtain a universal upper bound on the RHS of (17) that is independent of k, and thus an upper bound of $P_{\rm ph}$♯. This will be done in Section 5.

4. Upper confidence limit on the phase error rate $p_{\rm sft}(k,c)$

Now let us turn to the definition of $\hat{p}_{\rm sft}(c)$. As mentioned above, since length l of sample bits is finite in practical QKD systems, the phase error rate of a sifted key $p_{\rm sft}(k,c)$ deviates from that of sample bits, $p_{\rm smp}(c)$, due to statistical fluctuations. Hence, in order to guarantee the security by privacy amplification, instead of $p_{\rm smp}(c)$, one needs to use the estimated upper bound $\hat{p}_{\rm sft}(c)$ of $p_{\rm sft}(k,c)$, defined with the statistical effect taken into account.

As long as $p_{\rm sft}(k,c)$ is estimated larger than the actual value, i.e., $\hat{p}_{\rm sft}(c) > p_{\rm sft}(k,c)$, there is no loss of security, because then, more information is erased by the privacy amplification than is actually leaked to Eve. On the other hand, however, one needs to avoid a situation where $p_{\rm sft}(k,c)$ is estimated smaller as $\hat{p}_{\rm sft}(c) \le p_{\rm sft}(k,c)$. In such a case, the privacy amplification of the previous section does not work since $[g(k,c)]^- = 0$. Hence, at least as a necessary condition, the function $\hat{p}_{\rm sft}$ needs to satisfy that

$$\Pr_k \{\ c \mid \hat{p}_{\rm sft}(c) \ge p_{\rm sft}(k,c)\ \} > 1-\varepsilon \quad \text{for } \forall k, \tag{18}$$

where $\Pr_k\{c|Q\}$ denotes the probability that c occurs satisfying a condition Q, under the hypergeometric distribution $P_{\rm hg}(c|k)$. In order to maximize the key generation rate for fixed values of l, n, we wish to minimize $\hat{p}_{\rm sft}(c)$ as small as possible. In statistics, this corresponds to an interval estimation problem. That is, finding $\hat{p}_{\rm sft}(c)$ satisfying (18) is to obtain an upper confidence limit on $p_{\rm sft}(k,c)$ from an observed value of c, with significance level ε (see, e.g., [27]).

In the following, we derive the minimum estimate $\hat{p}_{{\rm sft},\varepsilon}(c) = \hat{p}_{\rm sft}(c)$ satisfying the condition (18) under the normal approximation of $P_{\rm hg}(c|k)$ by employing interval estimation of k. Although there is a standard procedure found in every textbook for this analysis (e.g., [27]), we reproduce it below for the sake of explanation. First we define the normal distribution function by

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-y^2/2) dy, \tag{19}$$

and $s(\varepsilon)$ as the deviation corresponding to ε, e.g.,

$$s(\varepsilon) = \Phi^{-1}(\varepsilon) \tag{20}$$

♯ A similar analysis was given by Fung et al. [26]. However, they seem to evaluate $P_{\rm hg}(c|k)S_{\rm pa}(k,c)$ without the summation. This corresponds to the probability that a certain set of values k and c occur *and then* the virtual phase error correction by Alice and Bob fails.

such that $\varepsilon = \Phi(s(\varepsilon))$. In what follows, we often abbreviate $s(\varepsilon)$ to s. Then, by applying the normal approximation to $P_{\rm hg}(c|k)$, we have the relation

$$\Pr_k \{ \ c \ | \ c \geq \bar{c}(k) - s(\varepsilon)\sigma(k) \ \} > 1 - \varepsilon \tag{21}$$

for any integer k; that is, $c \geq \bar{c}(k) - s(\varepsilon)\sigma(k)$ holds at least with probability $1-\varepsilon$ for any integer k. Note that this condition is equivalent to $(c - \bar{c}(k))^2 \leq s(\varepsilon)^2\sigma(k)^2$ or $c \geq \bar{c}(k)$. We rewrite this condition further as

$$(p_{\rm smp} - p)^2 \leq 4\gamma p(1-p), \text{ or } p_{\rm smp} \geq p \tag{22}$$

where $p = k/(n+l)$, $p_{\rm smp}(c) = c/l$, and

$$\gamma := \frac{s(\varepsilon)^2 n}{4l(n+l-1)}. \tag{23}$$

The condition (22) is equivalent to $p \leq \hat{p}_\varepsilon(c)$, where $\hat{p}_\varepsilon(c)$ is a solution of $(p_{\rm smp} - \hat{p}_\varepsilon)^2 = 4\gamma\hat{p}_\varepsilon(1-\hat{p}_\varepsilon)$ given by

$$\hat{p}_\varepsilon(c) := \frac{1}{1+4\gamma}\left(p_{\rm smp} + 2\gamma + 2\sqrt{\gamma\left\{p_{\rm smp}\left(1-p_{\rm smp}\right) + \gamma\right\}}\right). \tag{24}$$

That is, $k/(n+l) = p \leq \hat{p}_\varepsilon(c)$ holds at least with probability $1-\varepsilon$ for any integer k. In other words, the rate $\hat{p}_\varepsilon(c)$ gives the upper bound of one-sided interval estimation of $p = k/(n+l)$. Using this estimate, we define another function

$$\hat{p}_{{\rm sft},\varepsilon}(c) := (\hat{p}_\varepsilon(c)(n+l) - c)/n = \frac{(n+l)\hat{p}_\varepsilon(c) - l p_{\rm smp}(c)}{n}. \tag{25}$$

Then, again, the inequality $\hat{p}_{{\rm sft},\varepsilon}(c) \geq p_{\rm sft}(k,c) = (k-c)/n$ holds at least with probability $1-\varepsilon$ for any integer k. As a result, by choosing $\hat{p}_{\rm sft}(c)$ as $\hat{p}_{{\rm sft},\varepsilon}(c)$, we can satisfy the condition (18). Throughout the paper, we will use these definitions of $\hat{p}_\varepsilon(c)$ and $\hat{p}_{{\rm sft},\varepsilon}(c)$ in calculating $\alpha(c)$.

Now two remarks are in order. First, if there are sufficiently many samples (i.e., with l large and thus γ sufficiently small), the error number c has roughly the same distribution, irrespective of whether the samples are picked up with or without replacement. In such a case, as we mentioned under Eq. (12), the hypergeometric distribution $P_{\rm hg}(c|k)$ can be approximated by the binomial distribution. Indeed, to the first order of $\sqrt{\gamma}$, the estimated value $\hat{p}_\varepsilon(c)$ of Eq. (24) can be approximated as

$$\begin{aligned}\hat{p}_\varepsilon(c) &\simeq p_{\rm smp}(c) + \frac{s}{l}\sqrt{\frac{n}{n+l-1}}\,\sigma_{\rm bin}(c) \\ &= p_{\rm smp}(c) + \frac{s}{l}\sqrt{\frac{n}{n+l-1}}\sqrt{l p_{\rm smp}(c)(1-p_{\rm smp}(c))},\end{aligned}$$

where $\sigma_{\rm bin}(c) := \sqrt{l p_{\rm smp}(c)(1-p_{\rm smp}(c))}$ denotes the deviation of the binomial distribution with the error rate of the sample bits being $p_{\rm smp}(c) = c/l$. Furthermore, by using the inequality $p_{\rm smp}(c) + \frac{s}{l}\sqrt{\frac{n}{n+l-1}}\,\sigma_{\rm bin}(c) \leq p_{\rm smp}(c) + \frac{s}{l}\sigma_{\rm bin}(c)$, and by noting that the larger $\hat{p}_\varepsilon(c)$ always gives better a security bound, we can instead use a simpler approximation given by

$$\hat{p}_\varepsilon(c) \simeq p_{\rm smp}(c) + \frac{s}{l}\sigma_{\rm bin}(c), \tag{26}$$

The approximated upper bound of (26) can also be obtained by an argument similar to the above, with the hypergeometric distribution replaced by the binomial distribution. This means that, for l sufficiently large, one can conclude that the phase error rate $p(k,c)$ of the total bits can be bounded from above by $\hat{p}_\varepsilon(c)$ of (26), which is simply the measured error rate $p_{\rm smp}(c)$ of the samples, plus s times its standard deviation $\frac{s}{l}\sigma_{\rm bin}$. The actual value deviates this bound only with a probability less than $\Phi(s)$; or in other words, this estimation fails only with a probability less than $\Phi(s)$.

5. Upper bounds on the decoding error probability $P_{\rm ph}$

Throughout the paper, we assume that Alice and Bob perform the protocol specified in Section 2, using the estimated upper bound $\hat{p}_{{\rm sft},\varepsilon}(c)$ of (24) and (25), obtained in the previous section. That is, we here substitute $\hat{p}_{{\rm sft},\varepsilon}(c)$ for $\hat{p}_{\rm sft}(c)$ in (1), and as a result of that, Alice and Bob use sacrifice bits of $\alpha(c) = h\left(\hat{p}_{{\rm sft},\varepsilon}(\max[c, c_{\rm min}])\right) + D$ in the PA step. In this setting, we evaluate the decoding error probability evaluate $P_{\rm ph}$ and obtain several upper bounds.

5.1. The Straightforward Upper Bounds

In Section 3.3, we showed that, in order to bound $P_{\rm ph}$, it suffices to bound $S_{\rm av}(k)$ of (17) for all values of k. In this subsection, we first present a simple evaluation of $P_{\rm ph}$, where we divide the summation $S_{\rm av}(k)$, given in (17), into two regions of c. This method is similar to those used in preceding literature [2, 3], and we call it here the *straightforward method*.

For each value of k, we set the boundary value $c_{\rm bnd}(k) := \lfloor \bar{c}(k) - s\sigma(k) \rfloor$, and divide the summation of (17) as

$$S_{\rm av}(k) = \sum_{c=0}^{c_{\max}} P_{\rm hg}(c|k) S_{\rm pa}(k,c) \tag{27}$$

$$\leq \sum_{c=0}^{\lfloor \bar{c}(k)-s\sigma(k) \rfloor} P_{\rm hg}(c|k) + \sum_{c=\lfloor \bar{c}(k)-s\sigma(k) \rfloor+1}^{c_{\max}} P_{\rm hg}(c|k) S_{\rm pa}(k,c) \tag{28}$$

$$\leq \sum_{c=0}^{\lfloor \bar{c}(k)-s\sigma(k) \rfloor} P_{\rm hg}(c|k) + \max_{c\in[\bar{c}(k)-s\sigma(k), c_{\max}]} S_{\rm pa}(k,c). \tag{29}$$

(In what follows, we often write $\bar{c}$, σ, s instead of $\bar{c}(k)$, $\sigma(k)$, $s(\varepsilon)$.) Then, by using the properties of $\hat{p}_{{\rm sft},\varepsilon}(c)$ given in the preceding section, the two terms of (29) can be evaluated as follows:

(i) The first summation of (29) is the probability $\Pr_k\{\ c \mid c < \bar{c}(k) - s(\varepsilon)\sigma(k)\ \}$. As we have shown in the preceding section, this term is less than ε (see (21)), if one applies the normal approximation to $P_{\rm hg}(c|k)$. To put it more explicitly, apply the

normal approximation of the form:

$$\sum_{c=a}^{b} P_{\rm hg}(c|k) \simeq \frac{1}{\sqrt{2\pi}} \int_{\zeta_a}^{\zeta_b} e^{-x/2} dx \tag{30}$$

with $\zeta_c := (c - \bar{c}(k))/\sigma(k)$. Then it follows that the first term of (29) is less than $\Phi(s(\varepsilon)) = \varepsilon$, where $\Phi(s)$ is the normal distribution function given in (19).

(ii) In the second term of (29), the function $S_{\rm pa}(k,c) = 2^{[g(k,c)]^- + 1}$ is maximized at $c = \bar{c}(k) - s\sigma(k)$, because $g(k,c)$, defined in (14), is decreasing with c. Also note that

$$\hat{p}_{{\rm sft},\varepsilon}(\bar{c}(k) - s\sigma(k)) = p_{\rm sft}(k, \bar{c}(k) - s\sigma(k))$$

holds by the definition of $\hat{p}_{{\rm sft},\varepsilon}(c)$, given in (24) and (25).†† Thus from (14), we have

$$\begin{aligned} &g\left(k, \bar{c}(k) - s\sigma(k)\right) = nh\left(p_{\rm sft}(k, \bar{c}(k) - s\sigma(k))\right) - \alpha\left(\bar{c}(k) - s\sigma(k)\right) \\ &\leq nh\left(p_{\rm sft}(k, \bar{c}(k) - s\sigma(k))\right) - nh\left(\hat{p}_{{\rm sft},\varepsilon}\left(\bar{c}(k) - s\sigma(k)\right)\right) - D = -D \end{aligned}$$

For the inequality of the second line, we used the fact that $\alpha(c) = h\left(\hat{p}_{{\rm sft},\varepsilon}(\max[c, c_{\min}] + 2)\right) \geq h\left(\hat{p}_{{\rm sft},\varepsilon}(c)\right)$. This means that the second summation of (29) can be bounded by 2^{-D+1}. We remark that, unlike the first term of (29), this upper bound is valid without relying on the normal approximation.

Note here that the both bounds are valid for all values of k. Hence by combining these two upper bounds, we obtain the following proposition.

Proposition 1 *For a given ε (and the corresponding $s(\varepsilon) = \Phi^{-1}(\varepsilon)$), suppose that $c_{\min} \leq c_{\max}$, and that Alice and Bob perform the QKD protocol specified in Section 2. Then by applying the normal approximation to $P_{\rm hg}(c|k)$, $P_{\rm ph}$ can be bounded as*

$$P_{\rm ph} \leq \max_k S_{\rm av}(k) \leq \varepsilon + 2^{-D+1}. \tag{31}$$

If one wishes to bound $P_{\rm ph}$ by a certain value, say $P_{\max}$, a convenient choice of parameters is $\varepsilon = 2^{-D+1} = \frac{1}{2}P_{\max}$, or equivalently, $D = 2 - \log_2 P_{\max}$ and $s = \Phi^{-1}(\varepsilon) = \Phi^{-1}\left(\frac{1}{2}P_{\max}\right)$.† Then Inequality (10) guarantees that the trace distance is bounded as $\|\rho_{A,E'} - \rho_{\rm Ideal}\|_1 \leq 2\sqrt{2}\sqrt{P_{\max}}$, if Alice and Bob use a universal$_2$ hash function family that consists of linear and surjective functions.

Further, if parameters l and n are sufficiently large, we can also obtain a tight bound on the first term of (29) *without* relying on the normal approximation of $P_{\rm hg}(c|k)$.

††In fact, this is exactly the way we planned when we defined $\hat{p}_{{\rm sft},\varepsilon}(c)$: As mentioned in sentences below (46), the function $\hat{p}_\varepsilon(c)$ is defined so that the condition $\hat{p}_\varepsilon(\bar{c}(k) - s\sigma(k)) = p(k)$ is satisfied for all k. This condition is equivalent to $\hat{p}_{{\rm sft},\varepsilon}(\bar{c}(k) - s\sigma(k)) = p_{\rm sft}(k, \bar{c}(k) - s\sigma(k))$, due to definitions of $\hat{p}_{{\rm sft},\varepsilon}(c)$ and $p_{\rm sft}(k,c)$ given in (25) and in Table 1.

† Of course, the optimal choice is to let $\varepsilon = aP_{\max}$ and $2^{-D+1} = (1-a)P_{\max}$, and then find the optimal $0 < a < 1$ that yields the largest key generation rate. However, we do not pursue this optimality in the rest of the paper, since varying a contributes very little to the key rate in typical situations.

Lemma 1 *If* $\frac{5}{4}s(\varepsilon)^2 \le l \le n$, $1 \le k$, *and* $c_{\max} \le 0.12l$, *we have*

$$\sum_{c=0}^{\min(\lfloor \bar{c}-s\sigma \rfloor, c_{\max})} P_{\rm hg}(c|k) \le \sqrt{\frac{n+l}{n}}\sqrt{\frac{s(\varepsilon)^2+2\pi}{2}}\, e^{\mu}\varepsilon, \tag{32}$$

where $\mu := 1/(6n) + 1/(12)$. *Note that this bound holds rigorously, without relying on the normal approximation of* $P_{\rm hg}(c|k)$.

This lemma will be proved in Appendix B.3.

Now recall that the upper bound 2^{-D+1}, obtained above for the second term of (29), does not rely on any approximation either. Hence, besides Proposition 1, we can obtain another bound on $P_{\rm ph}$ that is similarly tight, and is valid rigorously *without* relying on any approximation:

Proposition 2 *Suppose that* $\frac{5}{4}s(\varepsilon)^2 \le l \le n$, *and* $c_{\max} \le 0.12l$ *are satisfied for a given* ε *(i.e., with* $\Phi(s) = \varepsilon$*). Also assume that Alice and Bob perform the QKD protocol specified in Section 2. Then without using the normal approximation of* $P_{\rm hg}(c|k)$, *we have*

$$P_{\rm ph} \le \max_k S_{\rm av}(k) \le \sqrt{\frac{s(\varepsilon)^2+2\pi}{2}}\sqrt{\frac{n+l}{n}}\, e^{\mu}\varepsilon + 2^{-D+1}. \tag{33}$$

5.2. The Upper Bounds by The Gaussian Integration

In the above analysis of the straightforward bounds, if one wishes to bound $P_{\rm ph}$ by a certain value, say $P_{\max}$, it is necessary to let $D \ge 1 - \log_2 P_{\max}$. Hence, if one choose a very small $P_{\max}$ in order to achieve a high level security, this D can decrease the final key length severely through the sacrificed bit length (1).

In this subsection, we derive improved bounds that holds with $D = 1$. We call them here the *Gaussian bounds* for the following reason. The first step of the analysis is similar to that of the previous section; i.e., we divide the summation of $S_{\rm av}(k)$ as in (28) and obtain upper bounds for each term. For the first term of (28), we use the normal approximation (30) again and bound it by ε. However, for the second term of (28), we employ a quite different strategy: We approximate $P_{\rm hg}(k|c)$ by using (30), and also upper bound $S_{\rm pa}(k,c)$ by an exponential function of a simple linear function of c (specified below in (35)). By using this simple form, we evaluate the summation over c as a Gaussian integral. As a result of this integration, instead of 2^{-D+1} appearing in the previous subsection, we obtain an upper bound $\delta\varepsilon$ on the second term, with δ being small for large l, n.

In order for this strategy using the Gaussian integration to work properly, parameter k must be confined to a specific region. Thus as a preparation, we consider the following three cases depending on the value of k:

(i) If k is too small (i.e., $0 \le k \le nc_{\min}/l$), it can be shown that $S_{\rm pa}(k,c)$ is always bounded by ε, by using the properties of $g(k,c)$. Thus $S_{\rm av}(k) \le \varepsilon$.

(ii) For the intermediate domain where $nc_{\min}/l \le k \le (n+l)\hat{p}_{\mathrm{sft},\varepsilon}(c_{\max})$, the function $g(k,c)$ (used for $S_{\mathrm{pa}}(k,c) = 2^{[g(k,c)]^- + 1}$) can be bounded from above by a simple function, i.e., a constant or a linear function of c.

(iii) If k is too large (i.e., $(n+l)\hat{p}_{\mathrm{sft},\varepsilon}(c_{\max}) \le k$), we can also show that $S_{\mathrm{av}}(k)$ is less than $\sum_{c=0}^{\bar{c}-s\sigma} P_{\mathrm{hg}}(c|k)$.

The more precise argument will be given in Appendix C, and we have the following theorem.

Theorem 1 *Let $D = 1$. If $c_{\min} \le c_{\max}$ and $2 \le s(\varepsilon)$, then $S_{\mathrm{av}}(k)$ is bounded from above as follows*

- *(Case 1) If $0 \le k \le nc_{\min}/l$,*

$$S_{\mathrm{av}}(k) \le \varepsilon. \tag{34}$$

- *(Case 2) If $nc_{\min}/l < k \le (n+l)\hat{p}_{\mathrm{sft},\varepsilon}(c_{\max})$, for an arbitrary possible outcome c, we have*

$$S_{\mathrm{pa}}(k,c) \le \min\left(2^{-\beta(c-(\bar{c}-s\sigma+1))},\ 1\right), \tag{35}$$

where

$$\beta := \frac{1}{1+4\gamma}\frac{n+l}{l} h'(\hat{p}_{\mathrm{sft},\varepsilon}(c_{\max})). \tag{36}$$

Thus

$$S_{\mathrm{av}}(k) \le \sum_{c=0}^{\min(\lfloor \bar{c}-s\sigma \rfloor, c_{\max})} P_{\mathrm{hg}}(c|k) + \sum_{c=\lfloor \bar{c}-s\sigma \rfloor+1}^{c_{\max}} P_{\mathrm{hg}}(c|k) 2^{-\beta(c-(\bar{c}-s\sigma)+1)}. \tag{37}$$

- *(Case 3) If $(n+l)\hat{p}_{\mathrm{sft},\varepsilon}(c_{\max}) \le k$, then $c_{\max} \le \bar{c} - s\sigma$ holds by the definition of $\hat{p}_{\mathrm{sft},\varepsilon}(c)$. Hence*

$$S_{\mathrm{av}}(k) \le \sum_{c=0}^{c_{\max}} P_{\mathrm{hg}}(c|k) \le \sum_{c=0}^{\lfloor \bar{c}-s\sigma \rfloor} P_{\mathrm{hg}}(c|k). \tag{38}$$

(For the proof of this theorem, see Appendix C.) We stress that the normal approximation to $P_{\mathrm{hg}}(c|k)$ is not yet applied, and thus all inequalities are rigorous at this stage‡

Then in the rest of this subsection, we will show that the right hand side of each inequality of Theorem 1 can be bounded from above by $(1+\delta)\varepsilon$, with δ being smaller than one for sufficiently large l, n. In other words, we obtain an upper bound on $S_{\mathrm{av}}(k)$ that is valid for all k; and thus an upper bound on P_{ph} (recall the argument of Section 3.3). can be bounded from above by (and thus P_{ph}) from above by ε. Let us first discuss

‡ It is true that we used the normal approximation in deriving $\hat{p}_{\mathrm{sft},\varepsilon}(c)$ in (25) and (24), and that $\hat{p}_{\mathrm{sft},\varepsilon}(c)$ is used in the statement of Theorem 1. However, in the proof of Theorem 1 we use no approximation; thus the theorem holds rigorously, without any approximation.

the easier cases, namely, Cases 1 and 3. As mentioned above, for these two cases $S_{\mathrm{av}}(k)$ can be easily shown to be less than ε: For Case 1, it is already proved in Theorem 1. For Case 3, if one applies the normal approximation to $P_{\mathrm{hg}}(c|k)$, $S_{\mathrm{av}}(k)$ is bounded by ε, as can be seen by the same argument as in the previous section (see the paragraph of (30)).

Hence it remains to evaluate Case 2, where parameter k is restricted as $nc_{\min}/l < k \leq (n+l)\hat{p}_{\mathrm{sft},\varepsilon}(c_{\max})$. As mentioned above, we here show that $S_{\mathrm{av}}(k)$ can be rewritten as the Gaussian integration in this case. In Inequality (37), the first term on the right hand side can be bounded by ε, with the approximation applied to $P_{\mathrm{hg}}(c|k)$. For the second term, which is a summation over c, we replace $P_{\mathrm{hg}}(c|k)$ with the the normal approximation. In addition to that, we replace $S_{\mathrm{pa}}(k,c)$ appearing in the same summation by the right hand side of (35). Then the summation can be rewritten a Gaussian integral:

$$\sum_{c=\lfloor \bar{c}-s\sigma \rfloor}^{c_{\max}} P_{\mathrm{hg}}(c|k) 2^{-\beta(c-(\bar{c}-s\sigma)+1)} \tag{39}$$

$$\simeq \frac{1}{\sqrt{2\pi}} \int_{-s}^{(c_{\max}-\bar{c})/\sigma} \exp\left[-\frac{x^2}{2} - s\,(x+s)\,\xi_\varepsilon(k)\right] dx.$$

$$\leq \frac{1}{\sqrt{2\pi}} \int_{-s}^{\infty} \exp\left[-\frac{x^2}{2} - s\,(x+s)\,\xi_\varepsilon(k)\right] dx. \tag{40}$$

$$= e^{\frac{1}{2}\xi_\varepsilon(\xi_\varepsilon-2)s^2} \frac{1}{\sqrt{2\pi}} \int_{(\xi_\varepsilon-1)s}^{\infty} e^{-x^2/2} dx \tag{41}$$

$$=: I_2\left(\xi_\varepsilon(k)\right),$$

where

$$\xi_\varepsilon(k) := (\ln 2)\beta\sigma(k)/s(\varepsilon).$$

Further, in order to bound $I_2\left(\xi_\varepsilon(k)\right)$ using ε, we introduce the inequalities

$$\frac{\sqrt{2}}{\sqrt{x^2+2\pi}} e^{-x^2/2} \leq \Phi(x) \leq \frac{\sqrt{2}}{x} e^{-x^2/2}, \tag{42}$$

where $\Phi(x)$ is the normal distribution function given in (19). (Inequalities (42) will also be proved in Appendix C.) By using (42), the integral $I_2\left(\xi_\varepsilon(k)\right)$ can be evaluated further as

$$I_2\left(\xi_\varepsilon(k)\right) \leq \frac{\sqrt{1+2\pi s^{-2}}}{\xi_\varepsilon(k)-1} \Phi(s(\varepsilon)) = \frac{\sqrt{1+2\pi s^{-2}}}{\xi_\varepsilon(k)-1}\,\varepsilon. \tag{43}$$

Note here that $\sigma(k)$ is an increasing function of k, because $\xi_\varepsilon(k)$ is. Thus the final term of (43) is maximized at the lower boundary $k = nc_{\min}/l$, and we obtain finally

$$I_2\left(\xi_\varepsilon(k)\right) \leq \frac{\sqrt{1+2\pi s^{-2}}}{\xi_{\min,\varepsilon}-1}\,\varepsilon \tag{44}$$

with $\xi_{\min,\varepsilon} := \xi_\varepsilon(nc_{\min}/l)$. We now have the following theorem:

Theorem 2 *For a given ε, suppose that $c_{\min} \le c_{\max}$, $2 \le s(\varepsilon)$ and $1 < \xi_{\min,\varepsilon}$ with*

$$\begin{aligned}\xi_{\min,\varepsilon} &:= \xi_\varepsilon(nc_{\min}/l) \\ &= \frac{(n+l)\ln 2}{s(\varepsilon)l(1+4\gamma)} h'\left(\hat{p}_{\mathrm{sft},\varepsilon}(c_{\max})\right)\sigma(nc_{\min}/l).\end{aligned} \tag{45}$$

Here $\hat{p}_{\mathrm{sft},\varepsilon}(c)$ is defined in Eq. (25), σ in Eq. (12), and $h'(x) = \log_2\left(\frac{1-x}{x}\right)$. Also assume that Alice and Bob perform the QKD protocol specified in Section 2. Then with the normal approximation applied to $P_{\mathrm{hg}}(c|k)$, P_{ph} can be bounded as

$$P_{\mathrm{ph}} \le \max_k S_{\mathrm{av}}(k) \le (1+\delta)\varepsilon, \tag{46}$$

where

$$\delta := \frac{\sqrt{1+2\pi s(\varepsilon)^{-2}}}{\xi_{\min,\varepsilon}-1}. \tag{47}$$

Note here that none of $c_{\min}$, $\hat{p}_{\mathrm{sft},\varepsilon}(c_{\max})$ or γ depends on k or c, which can vary for each run of the protocol; thus $\xi_{\min,\varepsilon}$ can be calculated as a fixed value specified by the protocol. (In other words, $\xi_{\min,\varepsilon}$ is the constant and thus calculated at the preparation stage prior to the protocol.)

Further, as we have done in the previous subsection, if parameters l and n are sufficiently large, we can also obtain a similarly good bound *without* relying on the normal approximation of $P_{\mathrm{hg}}(c|k)$ (in Eq. (30)). By using exact upper bounds on $P_{\mathrm{hg}}(c|k)$ including Lemma 1, we obtain the following theorem:

Theorem 3 *Suppose that $1 \le l \le n$, $s^2 \le c_{\min} \le c_{\max} \le 0.12l$, and $1 < \xi_{\min}$ are satisfied for a given ε. Also assume that Alice and Bob perform the QKD protocol specified in Section 2. Then without using the normal approximation of $P_{\mathrm{hg}}(c|k)$, we have*

$$P_{\mathrm{ph}} \le \max_k S_{\mathrm{av}}(k) \le P_{\mathrm{ph},\varepsilon}(c_{\min}, \xi_{\min,\varepsilon}), \tag{48}$$

where

$$\begin{aligned}P_{\mathrm{ph},\varepsilon}(c_{\min}, \xi_{\min,\varepsilon}) := &\sqrt{\frac{s(\varepsilon)^2+2\pi}{2}}\sqrt{\frac{n+l}{n}}e^{\mu}\varepsilon \\ &+ \left(\frac{\sqrt{1+2\pi s(\varepsilon)^{-2}}}{\xi_{\min,\varepsilon}-1}\frac{e^{\mu+\nu}}{\sqrt{1-\frac{s(\varepsilon)}{\sqrt{c_{\min}}}}} + \varepsilon\right)\varepsilon,\end{aligned} \tag{49}$$

where $\mu = 1/(6n) + 1/12$, $\nu = 1/(12l) + 1/(2(n+l-1))$.

The proof of this theorem is given in Appendix D.

5.3. Second Order Asymptotics

Now, we roughly estimate the relation between the sacrifice bit length and the upper bound $\max_k S_{\mathrm{av}}(k)$ of the phase error. For this purpose, we focus on the asymptotic expansion for the sacrifice bit. In the protocol discussed in the above, the sacrifice

bit length $\alpha(c)$ is $\lceil nh\left(\hat{p}_{\mathrm{sft},\varepsilon}(c+1)\right)\rceil + 2$ with $\hat{p}_{\mathrm{sft},\varepsilon}(c) = \frac{(n+l)\hat{p}_\varepsilon(c) - l p_{\mathrm{smp}}(c)}{n}$ and $\hat{p}_\varepsilon(c) := \frac{1}{1+4\gamma}\left(p_{\mathrm{smp}} + 2\gamma + 2\sqrt{\gamma\left\{p_{\mathrm{smp}}\left(1-p_{\mathrm{smp}}\right)+\gamma\right\}}\right)$. When the ratio l/n is t, we obtain the asymptotic expansion:

$$\lceil nh\left(\hat{p}_{\mathrm{sft},\varepsilon}(c+1)\right)\rceil + 2 = nh\left(p_{\mathrm{smp}}(c_{\min})\right) + \sqrt{n} g_t(p_{\mathrm{smp}}(c_{\min})) + o(\sqrt{n}), \quad (50)$$

where $g_t(x) := h'(x)\sqrt{\frac{x(1-x)(1+t)}{4t}} s(\varepsilon)$. When we use only the first term in the above expansion, the upper bound $\max_k S_{\mathrm{av}}(k)$ for the phase error converges to zero or one. The limit value zero or one cannot be used for the approximation for the upper bound $\max_k S_{\mathrm{av}}(k)$ because the real value of the upper bound $\max_k S_{\mathrm{av}}(k)$ takes a value between zero and one, which is different from zero or one.

However, when we use up to the second order $\sqrt{n}$ in the asymptotic expansion of $\alpha(c)$, the upper bound $\max_k S_{\mathrm{av}}(k)$ converges to a value between zero and one. In this case, we can use the limit for the approximation for the upper bound $\max_k S_{\mathrm{av}}(k)$. That is, by using the above asymptotic expansion, the virtual phase error can be abounded as the following way.

Theorem 4 *For a given ε, $p_{\min}$, and $p_{\max}$, we choose $c_{\min}$ and $c_{\max}$ as $p_{\min} l$ and $p_{\max} l$, and assume that $l/n = t$. Also suppose that Alice and Bob perform the QKD protocol specified in Section 2, except that the sacrifice bit length $\alpha(c)$ is less than $nh\left(p_{\mathrm{smp}}(c_{\min})\right) + \sqrt{n} g_t(p_{\mathrm{smp}}(c_{\min}))$ for $c \in [c_{\min}, c_{\max}]$. Then, the maximum $P_{\mathrm{ph},n,l}$ of $S_{\mathrm{av}}(k)$ with given n and t can be asymptotically characterized as*

$$\lim_{n\to\infty} \max_{l: l \ge tn} P_{\mathrm{ph},n,l} \le \varepsilon. \quad (51)$$

The proof will be given in Appendix E.

6. How to use the above formulas to evaluate the security of one's QKD system

In this section we summarize what we have proved so far, and then explain how one can use Proposition 1 or 2, or Theorem 2 or 3 to evaluate the security of one's QKD system.

6.1. Summary of Our Results

As discussed in Section 3, the standard quantitative measure of the security of QKD is the trace distance $\left\|\rho_{A,E'} - \rho_{\mathrm{Ideal}}\right\|_1$ between the actual state $\rho_{A,E'}$ and the ideal state ρ_{Ideal}, given in (3). Inequalities (9) and (10) claim that this trace distance can be bounded from above by the averaged decoding error probability P_{ph} of the virtual phase error correction. Throughout the paper, we are interested in bounding P_{ph} by using the Shor-Preskill's formalism. Also in Section 3, we have shown that in order to bound P_{ph} under an arbitrary attack by Eve, it suffices to bound the probability $\max_k S_{\mathrm{av}}(k)$, with $S_{\mathrm{av}}(k)$ defined in (17) (or equivalently, for all k, one needs to bound $S_{\mathrm{av}}(k)$ by a certain value). Here the function $S_{\mathrm{av}}(k)$ gives an upper bound on the failure probability $S_{\mathrm{pa}}(k,c)$ of the virtual phase error correction, averaged with respect to the

hypergeometric distribution $P_{\rm hg}(c|k)$. Our analyses of Sections 4 and 5 are devoted for obtaining an upper bound on $\max_k S_{\rm av}(k)$.

In Section 4, we determined the suitable functional form of the upper bound $\hat{p}_{\rm sft}(c)$ on the phase error rate $\hat{p}_{\rm sft}(k,c)$ of the sifted key, such that we can achieve high key generation rates and the high-level security simultaneously. The function $\hat{p}_{\rm sft}(c)$ is used for calculating the sacrifice bit length $\alpha(c)$ of Eq. (1), i.e., the number of bits that needs to be erased in privacy amplification (PA). This problem can be reduced to determining an upper bound on parameter k, or equivalently, that on the phase error rate $p_{\rm sft}(k,c)$ of a sifted key. For this purpose, we derived an upper bound $\hat{p}_{{\rm sft},\varepsilon}(c)$ of Eqs. (24) and (25) on $p_{\rm sft}(k,c)$, as a function of the measured error rate $p_{\rm smp}(c) = c/l$ of sample bits. We here used the standard method of interval estimation, and the upper bound $\hat{p}_{{\rm sft},\varepsilon}(c)$ is defined so that, for any value of k, the undesired case $p_{\rm sft}(k,c) > \hat{p}_{{\rm sft},\varepsilon}(c)$ occurs with a probability $\leq \varepsilon$ (see Eqs. (18) and (21)).

Then in Section 5, by using this $\hat{p}_{{\rm sft},\varepsilon}(c)$ and the corresponding sacrificed bit length $\alpha(c)$ given in (1), we obtained the upper bounds on $S_{\rm av}(k)$ that holds for all k. By the argument of the paragraph of (17), this means that we have given upper bounds on $P_{\rm ph}$. For the sake of simplicity, we first gave straightforward bounds in Proposition 1 (with the approximated values of the hypergeometric distribution $P_{\rm hg}(c|k)$) and Proposition 2 (without any approximation). Next we gave the other bounds exploiting the properties of the Gaussian integration, which yield larger final key length G for sufficiently large l, n; namely, Theorem 2 (with the approximated $P_{\rm hg}(c|k)$) and Theorem 3 (without any approximation).

6.2. How to Use The Straightforward Upper Bounds

6.2.1. The Straightforward Upper Bound With The Normal Approximation (How to Use Proposition 1) Here we present how to calculate the secret key length of one's QKD system using the straightforward upper bound on $P_{\rm ph}$ obtained in Propositions 1.

- Preparation steps:
  - (i) Determine one's desired upper bound $T_{\max}$ on trace distance.
  - (ii) Calculate the corresponding upper bound on the phase error rate by $P_{\max} = \frac{1}{8}(T_{\max})^2$.
  - (iii) Let the confidence limit be $\varepsilon = \frac{1}{2}P_{\max}$. Then calculate parameter $s = \Phi^{-1}(\varepsilon)$, as the inverse value of the normal distribution function $\Phi(x)$ (see the definitions of $\Phi(x)$ and $s(\varepsilon)$ given in (19), (20)).
  - (iv) Let $D = \lceil 2 - \log_2 P_{\max} \rceil$.
  - (v) Determine $c_{\min}$ and $c_{\max}$.
  - (vi) (Parameter check:) No parameter check is necessary for Proposition 1.

Under this setting of parameters, one can guarantee that $P_{\rm ph} \leq \varepsilon + 2^{-D+1} \leq P_{\max}$, by applying the normal approximation to $P_{\rm hg}(c|k)$ and by using Proposition 1. Then Inequality (10) guarantees that the trace distance is bounded as $\|\rho_{A,E'} - \rho_{\rm Ideal}\|_1 \leq$

$2\sqrt{2}\sqrt{P_{\max}} = T_{\max}$. (As specified below, we here assume that Alice and Bob use a universal$_2$ hash function family that consists of linear and surjective functions.)

- For each run of the protocol:

  (vii) Perform the protocol as specified in Section 2. In particular in the PA step, for the calculation of the length $\alpha(c)$ of (1), use $\hat{p}_{\mathrm{sft},\varepsilon}(c)$ defined in Eqs. (24) and (25), as well as parameters s and D obtained in the preparation steps above.† Then use a universal$_2$ hash function family that consists of linear and surjective functions, to convert the reconciled key to the secret key.

As noted in Section 2, as a result of this protocol, Alice and Bob obtain the final key of length $G = n_{\mathrm{rec}} - \alpha(c)$ with $\alpha(c)$ given in (1), and n_{rec} being the reconciled key length. If an error correcting code with efficiency f is used, we have $n_{\mathrm{rec}} = n(1 - fh(p_{\mathrm{bit}}))$, with p_{bit} being the bit error rate of the sifted key. Thus Alice and Bob obtain the final key of length G, given in (2).

6.2.2. The Straightforward Upper Bound Without Any Approximation (How to Use Proposition 2) By using Proposition 2, an exact upper bound on P_{ph} can be obtained, without relying on the normal approximation of $P_{\mathrm{hg}}(c|k)$. In this case all the steps are the same as those given in Section 6.2.1, except for Steps (iii) and (vi):

(iii') Choose parameter s such that

$$\sqrt{\frac{n+l}{n}}\sqrt{\frac{s^2+2\pi}{2}}\, e^{\mu}\Phi(s) \leq \frac{1}{2}P_{\max}$$

is satified, where $\mu = 1/(6n) + 1/12$.

(vi') (Parameter check:) Check that $\frac{5}{4}s^2 \leq l \leq n$ and $c_{\max} \leq 0.12l$ are satisfied. If not, set $T_{\max}$ smaller and restart from Step (i).

As a result of Step (iii'), we have $\varepsilon = \Phi(s(\varepsilon)) \leq s^{-1} \times \frac{1}{2}P_{\max}$. This means that, for a fixed value of $P_{\max}$, one needs to choose $\varepsilon = \Phi(s(\varepsilon))$ to be smaller than that obtained in Section 6.2.1, by a factor of s^{-1}. As a result, s also turns out to be larger, one ends up with a smaller final key length. Note, however, that such increment of s is negligible for sufficiently large s (e.g., for $s \geq 10$), because $\Phi(s)$ scales as $e^{-\frac{1}{2}s^2}$ and thus a very small increment of s compensates the factor of s^{-1} in front of $\frac{1}{2}P_{\max}$. Hence the decrement in the final key length is very small. We will demonstrate this fact in the next section by a numerical calculation in Section 7.3.

6.3. How to Use The Upper Bounds by The Gaussian Integration (How to Use Theorems 2 and 3)

As mentioned in Section 5.2, if parameters l and n are sufficiently large, we can set $D = 1$ and still obtain similarly tight bounds on P_{ph} as given in Theorems 2 and 3;

† Throughout this section, we neglect the deviation of l, n from their averages when the bases x, z are chosen with a constant probability, and assume that they are constant.

thereby we can improve the final key length G. For these cases too, we summarize how to calculate the secret key length of one's QKD system.

6.3.1. The Gaussian Bound With The Normal Approximation (How to Use The Bound of Theorem 2) For Theorem 2, the preparation steps are modified as follows:

- Preparation steps:
  - (i) Determine one's desired upper bound on trace distance $T_{\max}$.
  - (ii) Calculate the corresponding upper bound on the phase error rate by $P_{\max} = \frac{1}{8}(T_{\max})^2$.
  - (iii) Set the confidence limit ε to be slightly smaller than $P_{\max}$. (For example, if l, n are sufficiently large, $\varepsilon = 0.9P_{\rm ph}$ is usually sufficient.) Then calculate parameter $s = \Phi^{-1}(\varepsilon)$, as the inverse value of the normal distribution function $\Phi(x)$ given in (19).
  - (iv) Let $D = 1$.
  - (v) Determine $c_{\min}$ and $c_{\max}$, such that the conditions in the first sentence of Theorem 2 are all satisfied.
  - (vi) (Parameter Check:) Check if δ is small enough so that Inequality (46) is satisfied. If not, go back to Step (iii) and set ε smaller.

After these preparation steps, Alice and Bob run the protocol as in previous sections. That is, they run the protocol as specified in Step (vii) of Section 6.2.1.

6.3.2. The Gaussian Bound Without The Normal Approximation (How to Use The Bound of Theorem 3) As we have done for the case of the straightforward bounds, we also obtained in Theorem 3 the exact version of the Gaussian bound that does not rely on the normal approximation of $P_{\rm hg}(c|k)$. This theorem was derived using essentially the same idea as Theorem 2 and achieves a similarly tight bound, but it does not rely on any approximation.

For Theorem 3, the preparation steps are the same as Theorem 2 (i.e., the same as in Section 6.3.1), except for Steps (v) and (vi):

(v") Determine $c_{\min}$ and $c_{\max}$, such that the conditions in the first sentence of Theorem 3 are all satisfied.

(vi") (Parameter Check:) Check if δ' is small enough so that Inequality (49) is satisfied. If not, go back to Step (iii) and set ε smaller.

After these preparation steps, Alice and Bob run the protocol as in previous sections. That is, they run the protocol as specified in Step (vii) of Section 6.2.1.

6.4. Rough Estimate of The Key Rate and The Security Parameter

We note here that if l, n are sufficiently large, parameters γ and δ becomes sufficiently small, and the approximate evaluation of the key length G of (2) can be greatly simplified.

As one can see from Steps (i) and (ii) of Section 6.3, bounding $P_{\rm ph}$ is enough for the security. If δ is sufficiently small, then according to Theorem 2 (or or Step (iii) of Section 6.3), $P_{\rm ph}$ can be bounded approximately by ε, which determines the value of $\hat{p}_{{\rm sft},\varepsilon}(c)$ via Eqs. (24) and (25). Then as we discussed in the paragraph of Eq. (26), if γ is sufficiently small, $\hat{p}_{{\rm sft},\varepsilon}(c) = \frac{n+l}{n}\hat{p}_\varepsilon(c) - \frac{l}{n}p_{\rm smp}(c)$ can be approximated by using $\hat{p}_\varepsilon(c) \simeq p_{\rm smp}(c) + \frac{s}{l}\sigma_{\rm bin}(c)$.

As a result, if the conditions of the first sentence of Theorem 2 are satisfied for a given set of experimental parameters, and if γ and δ are sufficiently small, one has the following rough estimates. The trace distance is approximately bounded by the square root of ε as

$$
\begin{aligned}
&||\rho_{A,E} - \rho_{\rm Ideal}|| \le 2\sqrt{2}\sqrt{P_{\rm ph}},\\
&P_{\rm ph} \le (1+\delta)\varepsilon \simeq \varepsilon.
\end{aligned}
$$

Parameter s is chosen to be the deviation of the standard deviation, i.e., $s = \Phi^{-1}(\varepsilon)$. Then this s determines the final key length G as

$$
\begin{aligned}
G &\simeq n\left[1 - fh(p_{\rm bit}) - h\left(\hat{p}_{{\rm sft},\varepsilon}(c)\right)\right],\\
\hat{p}_{{\rm sft},\varepsilon}(c) &= \frac{n+l}{n}\hat{p}_\varepsilon(c) - \frac{l}{n}p_{\rm smp}(c),\\
p_{\rm smp}(c) &= c/l,\\
\hat{p}_\varepsilon(c) &\simeq p_{\rm smp}(c) + \frac{s}{l}\sigma_{\rm bin}(c)\\
&= p_{\rm smp}(c) + \frac{s}{l}\sqrt{lp_{\rm smp}(c)(1-p_{\rm smp}(c))}.
\end{aligned}
$$

We expect that these relation will be useful for experimentalists and theorists who wish to obtain a rough estimate of the key length with the finite size effect taken into account.

7. Numerical results.

We demonstrate the tightness of our bound with numerical results. We consider a quantum channel in the absence of eavesdropper, and assume that it can be described as a binary symmetric channel with quantum bit error rate (QBER).

7.1. Case 1: Basis Choice with Probability $q = \frac{1}{2}$

First, as a comparison to preceding literature [2, 4], we plot key rates for the case where Alice and Bob choose the x and the z bases with the equal probability. We present two types of evaluations given in Section 6; one is the analysis of Section 6.2.2 using the straightforward bound of Proposition 2, the other is that of Section 6.3.2 using the Gaussian bound of Theorem 3. Note that both these bounds are derived *without* using the normal approximation; thus the all key generation rates obtained in this subsection are rigorous.

We assume that Alice and Bob choose both the phase basis and the bit basis with probability $q = 1/2$, and thus $n = l = N/4$. We also assume that Alice

and Bob consume $r = 40$ bits of a previously shared secret key for exchanging the hash value, in order to guarantee that $\epsilon_{\rm cor} \leq 10^{-12}$ (in the following, these $r = 40$ bits will be subtracted from the final key length G). Then we choose $P_{\rm max}$ to be $P_{\rm max} = 0.98 \times \frac{1}{8} \times 10^{-20}$, so that the trace distance $\|\rho_{A,E'} - \rho_{\rm Ideal}\|_1$ is guaranteed to be less than $T_{\rm max} = 2\sqrt{2P_{\rm max}} = 0.99 \times 10^{-10}$. By these choices of parameters, we can guarantee $T_{\rm max} + \epsilon_{\rm cor} \leq 10^{-10}$, which is the same condition as used in Ref. [4].

Because $r = 40$ bits are consumed for guaranteeing that Alice's and Bob's final keys are equal, the effective final key length is $G(c) - r$, with $G(c)$ defined in (2). Hence in this section, we define the final key rate to be

$$\begin{aligned} R(c) &:= \frac{G(c) - r}{n} \\ &= \frac{1}{n}\left[n\left(1 - fh(c/l)\right) - \lceil nh\left(\hat{p}_{{\rm sft},\varepsilon}(\max\{c, c_{\min}\} + 2)\right)\rceil - (D + r)\right]. \end{aligned} \tag{52}$$

The efficiency of bit error correction is chosen to be $f = 1.1$.

7.1.1. The Straightforward Bound With the above choices of parameters, we perform the analysis of Section 6.2.2, and obtain the corresponding final key rate R. Here we restrict ourselves to the case where parameters l, n satisfy $125 \leq l = n$. Parameters $P_{\rm max}$ and $T_{\rm max}$ are already specified above. As to parameter s, we follow Step (iii') and let $s = 9.9$, so that

$$\sqrt{\frac{n+l}{n}}\sqrt{\frac{s^2 + 2\pi}{2}}\, e^{\mu}\Phi(s) \leq \sqrt{s^2 + 2\pi}\, e^{1/4}\Phi(s) \leq 1.1 \times 10^{-22} \leq \frac{1}{2}P_{\rm max}.$$

According to Step (iv), we choose $D = \lceil 2 - \log_2 P_{\rm max}\rceil = 79$; next according to Step (v), $c_{\min} = 0.01l$ and $c_{\max} = 0.12l$. It is easy to verify that all these parameters are compatible with the parameter checks of Step (vi').

Then we assume that Alice and Bob perform the BB84 protocol (i.e., Step (vii)), in the quantum channels with QBER $= 1\%, 2.5\%$, and 5%. The corresponding key rates $R(c)$ (with $c = l \times$ QBER) are shown in bold curves in Fig. 1, versus $n + l$.

7.1.2. The Gaussian Bound For the same choice of parameters $q, r, P_{\rm max}, D$, and for the same ratio of $c_{\max} = 0.12l$ with respect to l, we perform the analysis of Section 6.3.2. The remaining parameters to be fixed are s and $c_{\min}$; hence we here numerically calculate the pairs of s and $c_{\min}$ that gives the best key rate $R(c)$. That is, we first fix l and n, and then search for the pair of s and $c_{\min}$ that is compatible with the parameter check and gives the largest $R(c)$. (This corresponds to repeating Steps (iii) through (vi') of Section 6.3.2, by letting ε smaller each time, until the largest key length $G(c)$ is obtained.) The results are shown in thin curves in Fig. 1.

As one can see from Fig. 1, if QBER=5%, the Gaussian bound gives better key rate than the straightforward bound for all l, n. On the contrary, for smaller QBER (1% and 2.5%), the straightforward bound becomes better for $l, n \simeq 5000$.

The dots in Fig. 1 represents the key rates obtained by Tomamichel et al. [4] under the same condition. It can be clearly seen that our key rates R are better in all parameter

regions. For example, Fig. 1 gives $R = 0.19$ for QBER $= 5\%$ and $n + l = 10^4$, while Tomamichel et al. gave $R = 0$ in this region [4]. As $n+l$ becomes larger, R converge very fast to the asymptotic values; all three curves reach more than 80% of the asymptotic values at $n + l = 2 \times 10^5$.

In particular, as the key size becomes larger, R converge very fast to the asymptotic values, more than 80% of the asymptotic values at $n+l = 2\times 10^5$. As we have noted in Section 2, key distillation is quite practical even in this region. That is, the sizes of bit error correcting codes are independent of security, and thus Alice and Bob may perform bit error correction by dividing a sifted key of n bits to arbitrarily smaller blocks. As to privacy amplification, one can use the efficient algorithm for the multiplication of the (modified) Toeplitz matrix and a vector.

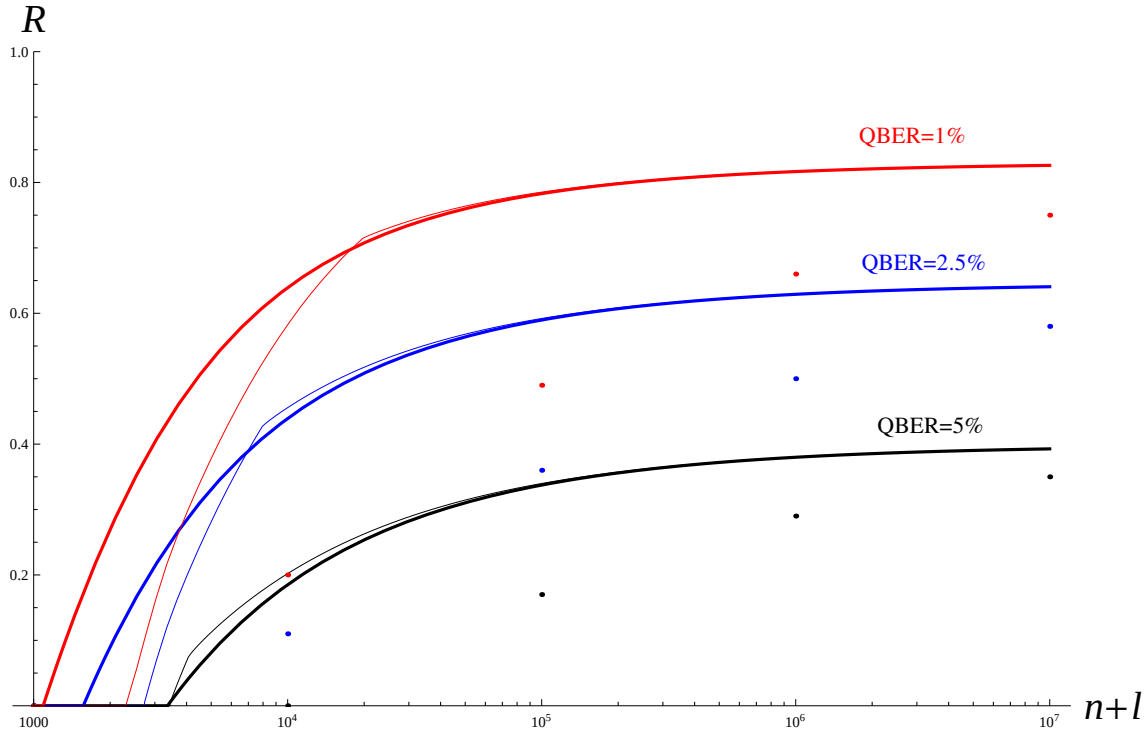

Figure 1. (Color online) Key generation rate $R = (G - r)/n$ versus $n + l$, which is the sum of lengths of a sifted key and sample bits. Here we assume that x and the z bases are chosen with the equal probability, i.e., $q = \frac{1}{2}$. The typical QBER are chosen to be 1% (red), 2.5% (blue), and 5% (black). As to the security, we set $r = 40$ and $P_{\max} < 0.98 \times \frac{1}{8} \times 10^{-20}$, so that $T_{\max} + \epsilon_{\rm corr} \le 10^{-10}$. That is, the sum of the trace distance and $\epsilon_{\rm cor}$ is less than 10^{-10}. We have used two types of analysis to achieve this value of $P_{\max}$: The bold curves represent the key rates based on the straightforward bound given in Proposition 2 and in Section 6.2.2. The thin curves are based on the Gaussian bound given in Theorem 3 and in Section 6.3.2. We stress that these curves are obtained *without* using the normal approximation. Dots of the same color are the rates obtained in Figure 2 of Ref. [4].

7.2. Case 2: Optimized Basis Choice with Variable Probability q

Next, as a more practical setting, we consider the case where Alice and Bob choose the x and the z bases with varying probabilities q, $1-q$ (thus, $l = q^2N$, $n = (1-q)^2N$). Then we maximize the secret fraction F, defined by

$$\begin{aligned} F(c) &= \frac{G(c) - r}{N} \\ &= \frac{1}{N}\left[n\left(1 - fh(c/l)\right) - \lceil nh\left(\hat{p}_{\mathrm{sft},\varepsilon}(\max\{c, c_{\min}\} + 2)\right)\rceil - (D + r)\right] \end{aligned} \tag{53}$$

with respect a fixed raw key length N, where G denotes the final key length. We use the analysis of Section 6.3.2 based on the Gaussian bound of Theorem 3 (without any approximation); hence again, all the final key rates obtained in this subsection are rigorous. We choose parameters $P_{\max}$, ϵ_{cor} are chosen to be the same as in the previous subsection. According to Step (iii), we let $s(\varepsilon) = 10.5$ so that $\varepsilon = 4.32 \times 10^{-26} \ll P_{\max}$. The channel error rates are chosen to be QBER = 1%, 2.5%, and 5%, respectively.

Under these settings, for each fixed value of N, we performed numerical simulations to select the optimal values of q and $c_{\min}$ that give the maximum value of $F(c)$. That is, we first fix N, and then search for the pair of q and $c_{\min}$ that is compatible with the parameter check of Step (vi") and gives the largest $F(c)$. The results are shown in Figure 2.

7.3. Exact Bounds Verses Approximate Bounds

All the key rates of the previous two subsections are rigorous, in the sense that they are obtained without using any approximation. In this final subsection, we demonstrate that, for practical parameter regions, the key rates are almost the same, whether one uses the analysis based on the normal approximation (i.e., Proposition 1 and Theorem 2), or those without any approximation (i.e., Proposition 2 and Theorem 3).

In Fig. 3, the solid curve shows $R(c)$ obtained in Section 7.1.1 with QBER=1%. On the other hand, the dashed curve in the same figure is the key rate $R(c)$ obtained for the same values of QBER and $P_{\max}, r, l, n$ by the procedure of Section 6.2.1; hence this curve is obtained by using Proposition 1, and thus relies on the normal approximation of P_{hg}. Similarly in Fig. 4, the solid curve shows $F(c)$ obtained in Section 7.1.2 with QBER=5%, whereas the dashed curve is obtained by using Theorem 2, which relies on the normal approximation (Here we performed the optimization of s and $c_{\min}$).

Note that for both of these cases, the exact key rate and approximate key rate are almost identical. These results suggest that the simple analysis using the normal approximation (i.e., Proposition 1 or Theorem 2) can be justified for the security evaluations of practical QKD systems.

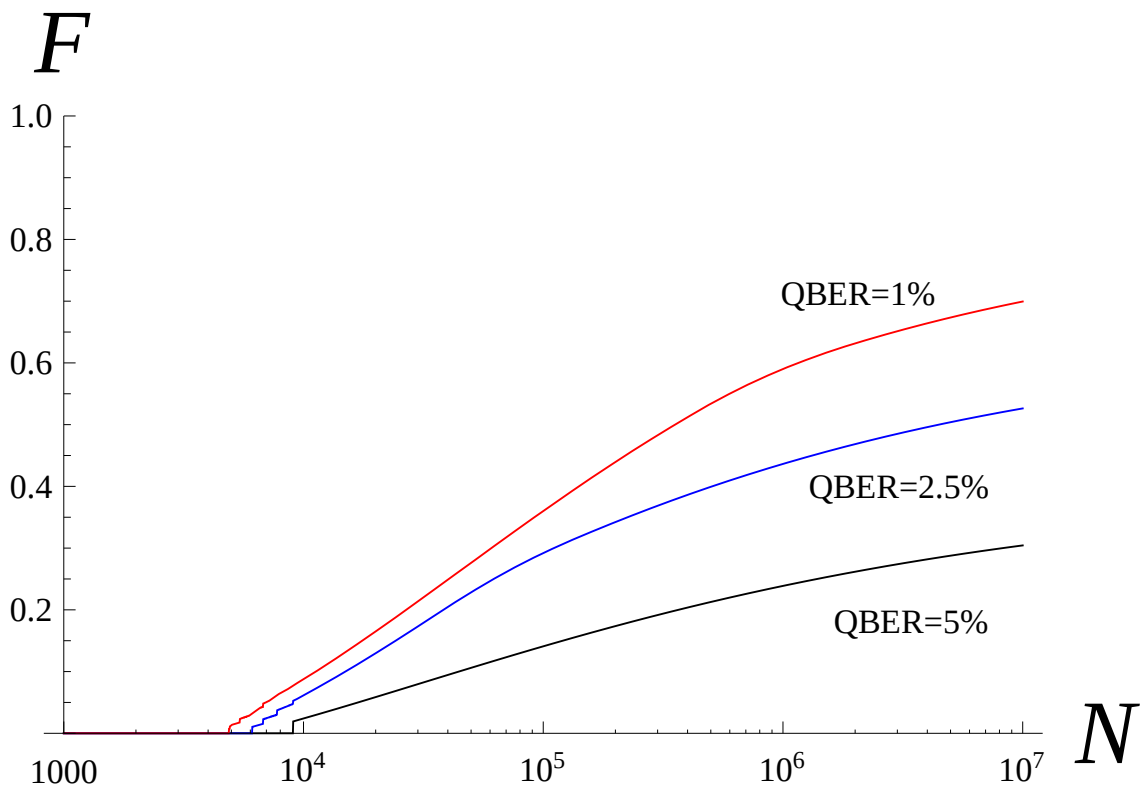

Figure 2. (Color online) Secret fraction $F = (G - r)/N$ versus raw key length N. Here we assume that Alice and Bob choose the x and the z bases with varying probabilities q, $1-q$. The probability q and the minimum errors $c_{\min}$ are also optimized to give maximum F. The typical QBER are chosen to be 1% (red), 2.5% (blue), and 5% (black). Parameters $P_{\rm ph}$, $\epsilon_{\rm cor}$ are chosen to be the same as in Figure 1, so that $T_{\max} + \epsilon_{\rm corr} \leq 10^{-10}$ is satisfied.

8. Summary

In this paper, we presented a concise analysis for the BB84 protocol that takes the finite key effect into account and yields better key generation rates, with and without relying on the normal approximation. Our results are indeed an improvement of preceding literature; as we have shown in Figure 1, our analysis give better key generation rates R in practical settings than in Refs. [2, 4].

In order to serve the convenience of experimentalists who wish to evaluate the security of their QKD systems, we included explicit procedures of security evaluation in Sections 3 and 6. In particular, in addition to presenting the exact values of key rates and security parameters, we also presented how to obtain their rough estimates using the normal approximation.

For the sake of simplicity, we restricted ourselves to the simple case where Alice has a perfect single photon source. On the other hand, in order to achieve a long communication distance by a practical QKD system using a weak coherent light source, decoy pulses are necessary [28]. This situation was analyzed by one of the authors [1],

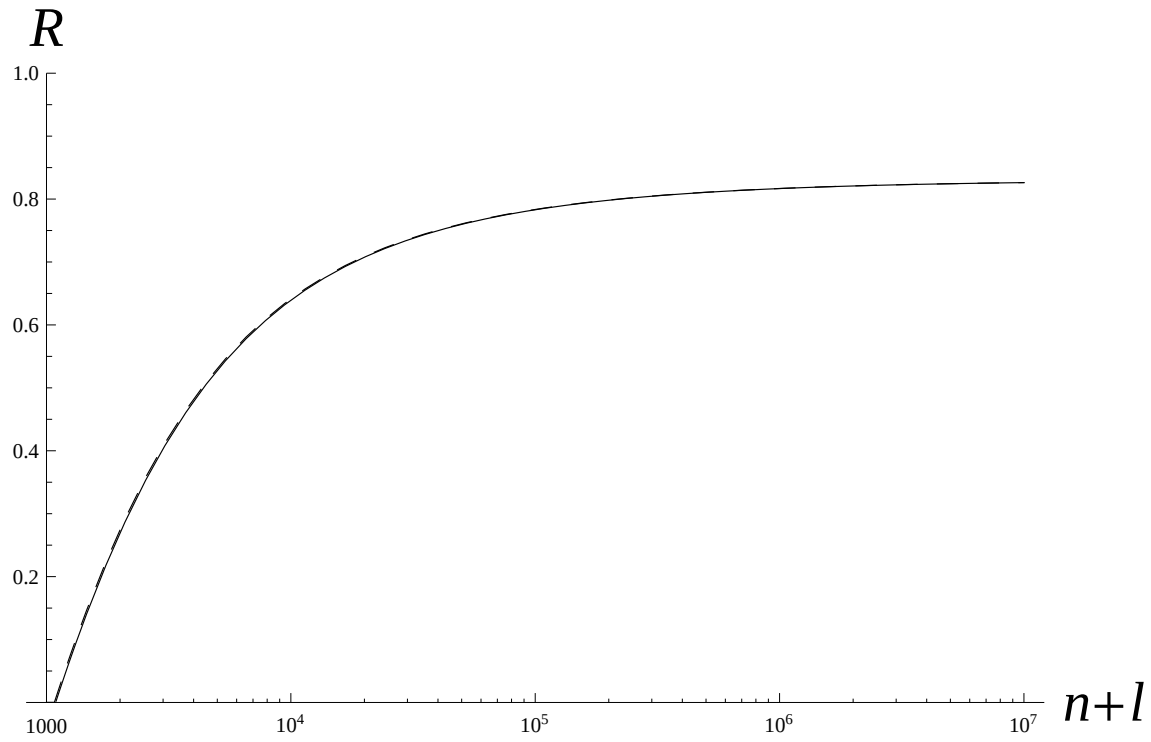

Figure 3. Solid Curve: the same curve as the solid curve in Figure 1 with QBER=1%. This curve is obtained by using Proposition 2, without using any approximation. Dashed Curve: The final key rate $R(c)$ obtained for the same values of QBER, $P_{\max}, r, l, n$, using the straightforward bounds of Proposition 1; hence this curve is obtained using the normal approximation. Note that the two curves are almost identical.

relying on the normal approximation. A thorough and exact analysis in this direction without any approximation remains as future work.

Acknowledgments The authors thank Ryutaroh Matsumoto for valuable comments. MH is partially supported by a MEXT Grant-in-Aid for Young Scientists (A) No. 20686026 and Grant-in-Aid for Scientific Research (A) No. 23246071. The Center for Quantum Technologies is funded by the Singapore Ministry of Education and the National Research Foundation as part of the Research Centres of Excellence programme. MH and TT are partially supported by the National Institute of Information and Communication Technolgy (NICT), Japan.

[1] M. Hayashi, Phys. Rev. A 76, 012329 (2007); ibid. 79, 019901(E) (2009).
[2] V. Scarani and R. Renner, Phys. Rev. Lett. 100, 200501 (2008).
[3] Y. Sano, R. Matsumoto, and T. Uyematsu, J. Phys. A 43, 495302 (2010).
[4] M. Tomamichel, C. C. W. Lim, N. Gisin, and R. Renner, arXiv:1103.4130v1 [quant-ph].
[5] H. Maassen, and J. B. M. Uffink, Phys. Rev. Lett. 60, 1103 (1988).
[6] J. M. Renes and J. -C. Boileau, Phys. Rev. Lett. 103, 020402 (2009).
[7] C. H. Bennett and G. Brassard, in Proceeding of the IEEE International Conference on Computers, Systems, and Signal Processing, Bangalore, India (IEEE, New York, 1984), pp.175-179.

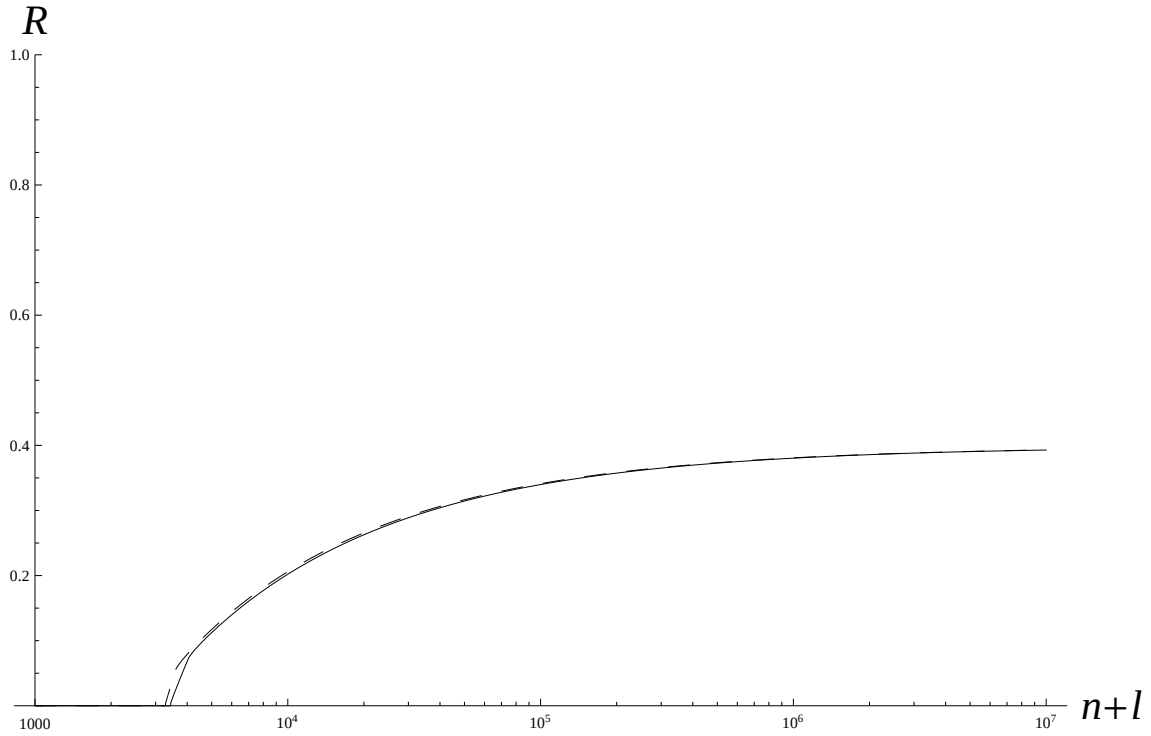

Figure 4. Solid Curve: the same curve as the thin curve in Figure 1 with QBER=5%. This curve is obtained by using Theorem 3 without using any approximation. Dashed Curve: The final key rate $R(c)$ obtained for the same values of QBER, $P_{\max}, r, l, n$, using the straightforward bounds of Theorem 2; hence this curve is obtained using the normal approximation. Note again that the two curves are almost identical.

[8] R. Renner, Security of Quantum Key Distribution, PhD thesis, Dipl. Phys. ETH, Switzerland, 2005; arXiv:quant-ph/0512258.

[9] R. Renner, and R. König, "Universally composable privacy amplification against quantum adversaries," Theory of Cryptography: Second Theory of Cryptography Conference, TCC 2005, J.Kilian (ed.) Springer Verlag 2005, vol. 3378 of Lecture Notes in Computer Science, pp. 407-425.

[10] H. -K. Lo and H. F. Chau, Science 283, 2050, (1999); P. W. Shor and J. Preskill, Phys. Rev. Lett. 85, 441 (2000).

[11] V. Strassen, "Asymptotische Abschätzugen in Shannon's Informationstheorie," In Transactions of the Third Prague Conference on Information Theory etc, 1962. Czechoslovak Academy of Sciences, Prague, pp. 689-723.

[12] M. Hayashi, "Information Spectrum Approach to Second-Order Coding Rate in Channel Coding," IEEE Transactions on Information Theory, Vol.55, No.11, 4947 - 4966 (2009);

[13] Y. Polyanskiy, H.V. Poor, S. Verdú, "Channel coding rate in the finite blocklength regime," IEEE Trans. Inform. Theory, Vol. 56, 2307 - 2359 (2010).

[14] M. B. Ruskai and E. Werner, eprint arXiv:math/9711207

[15] V. Chvátal, Discrete Mathematics 25, 285 (1979).

[16] M. Koashi, arXiv:quant-ph/0505108; New J. Phys. 11, 045018 (2009).

[17] T. Tsurumaru and M. Hayashi, arXiv:1101.0064v3 [quant-ph].

[18] T. Asai, and T. Tsurumaru, "Efficient Privacy Amplification Algorithms for Quantum Key Distribution" (in Japanese), IEICE technical report, ISEC2010-121 (2011).

[19] G. H. Golub, and C. F. Van Loan, Matrix Computation, Third Edition, (Johns Hopkins University Press, 1996).

[20] M. Ben-Or, Michal Horodecki, D. W. Leung, D. Mayers, J. Oppenheim "The Universal Composable Security of Quantum Key Distribution," Theory of Cryptography: Second Theory of Cryptography Conference, TCC 2005, J.Kilian (ed.) Springer Verlag 2005, vol. 3378 of Lecture Notes in Computer Science, pp. 386-406.
[21] S. Watanabe, R. Matsumoto, T. Uyematsu, and Y. Kawano, Phys. Rev. A 76, 032312 (2007).
[22] M. Hayashi, Phys. Rev. A 74, 022307 (2006).
[23] S. Watanabe, R. Matsumoto, and T. Uyematsu, International Journal of Quantum Information, vol. 4, no. 6, pp. 935-946, (2006).
[24] J. Håstad, R. Impagliazzo, L. A. Levin, and M. Luby, "A Pseudorandom Generator from any One-way Function," SIAM J. Comput. 28, 1364 (1999)
[25] C.H.Bennett, G. Brassard, C. Crepeau, and U.M. Maurer, "Generalized privacy amplification," *IEEE Trans. Inform. Theory*, vol. **41**, 1915–1923, 1995.
[26] C. -H. F. Fung, X. Ma, and H. F. Chau, Phys. Rev. A **81**, 012318 (2010).
[27] P. G. Hoel, Elementary Statistics, 4th Ed., (John Wiley & Sons, New York, 1969).
[28] W. -Y. Hwang, Phys. Rev. Lett. 91, 057901 (2003); H.-K. Lo, X. Ma, and K. Chen, Phys. Rev. Lett. 94, 230504 (2005); X. -B. Wang, Phys. Rev. Lett. 94, 230503 (2005).
[29] J. Justesen and T. Hoholdt, Course In Error Correcting Codes, European Mathematical Society (2004).
[30] S.N. Lahiri, A. Chatterjee, T. Maiti, "Normal approximation to the hypergeometric distribution in nonstandard cases and a sub-Gaussian Berry-Esseen theorem," Journal of Statistical Planning and Inference vol. 137, 3570 - 3590 (2007).

Appendix A. Justification for the restricting the argument to the generalized Pauli channel

The generalized Pauli channel is defined to be a channel where the phase error and the bit errors occur stochastically (i.e., with a classical probability). It is easy to see that, in this setting, the virtual phase error probability $P_{\rm ph}$ after the privacy amplification, mentioned in the main text, can clearly be defined. In Ref. [1], it is shown that the trace distance can be bounded from above by using $P_{\rm ph}$.

Here we demonstrate that, without loss of generality, this argument can be extended to the case where the quantum channel Λ between Alice and Bob is arbitrary and general. First, we consider the discrete twirling. For n-bits sequence $x = (x_1, \ldots, x_n)$ and $z = (x_1, \ldots, z_n)$, define the unitary matrix $U(x,z) := (X^{x_1} \otimes X^{x_2} \otimes \cdots \otimes X^{x_n})(Z^{z_1} \otimes Z^{z_2} \otimes \cdots \otimes Z^{z_n})$, where X is the bit flip operator and Z the phase flip operator. Then, the discrete twirling of Λ is defined as $\overline{\Lambda} := \sum_{\mathbf{z}} 2^{-2n} \Lambda_{\mathbf{z}}$, where $\mathbf{z} = (x,z)$ and $\Lambda_{x,z}(\rho) := U(x,z)\Lambda(U(x,z)\rho U(x,z)^\dagger)U(x,z)^\dagger$. In this paper, we treat the phase error and the bit error of the channel $\overline{\Lambda}$ due to the following reason.

Now, we denote the final state and the ideal state with the public information x by $\rho_{A,E'|x}(\Lambda)$ and $\rho_{{\rm Ideal}|x}(\Lambda)$ when the channel between Alice and Bob is Λ. Hence, our security criterion is $\sum_x P_{\rm pub}(x) \|\rho_{A,E'|x}(\Lambda) - \rho_{{\rm Ideal}|x}(\Lambda)\|_1$. Indeed, the distribution $P_{\rm pub}(x)$ depends on the channel Λ in general, however, it does not change even if the channel is replaced by $\Lambda_{\mathbf{z}}$ because the initial random variable is uniform and the hash function and error correction are linear. Also for the same reason, we have $\|\rho_{A,E'|x}(\Lambda) - \rho_{{\rm Ideal}|x}(\Lambda)\|_1 = \|\rho_{A,E'|x}(\Lambda_{\mathbf{z}}) - \rho_{{\rm Ideal}|x}(\Lambda_{\mathbf{z}})\|_1$. The state $\sum_{\mathbf{z}} 2^{-2n} \rho_{A,E'|x}(\Lambda_{\mathbf{z}}) \otimes |\mathbf{z}\rangle\langle\mathbf{z}|$ and $\sum_{\mathbf{z}} 2^{-2n} \rho_{{\rm Ideal}|x}(\Lambda_{\mathbf{z}}) \otimes |\mathbf{z}\rangle\langle\mathbf{z}|$ can be regarded as the state $\rho_{A,E'|x}(\overline{\Lambda})$ and $\rho_{{\rm Ideal}|x}(\overline{\Lambda})$ because

the classical information $\mathbf{z}$ can be treated as a part of Eve's system with the channel $\overline{\Lambda}$. Hence,

$$\begin{aligned}
&\sum_x P_{\rm pub}(x)\|\rho_{A,E'|x}(\Lambda)-\rho_{{\rm Ideal}|x}(\Lambda)\|_1 \\
=&\|\sum_{\mathbf{z}}2^{-2n}\sum_x P_{\rm pub}(x)\|\rho_{A,E'|x}(\Lambda_{\mathbf{z}})\otimes|\mathbf{z}\rangle\langle\mathbf{z}|-\rho_{{\rm Ideal}|x}(\Lambda_{\mathbf{z}})\otimes|\mathbf{z}\rangle\langle\mathbf{z}|\|_1 \\
=&\|\sum_x P_{\rm pub}(x)\|\sum_{\mathbf{z}}2^{-2n}\rho_{A,E'|x}(\Lambda_{\mathbf{z}})\otimes|\mathbf{z}\rangle\langle\mathbf{z}|-\sum_{\mathbf{z}}2^{-2n}\rho_{{\rm Ideal}|x}(\Lambda_{\mathbf{z}})\otimes|\mathbf{z}\rangle\langle\mathbf{z}|\|_1 \\
=&\sum_x P_{\rm pub}(x)\|\rho_{A,E'|x}(\overline{\Lambda})-\rho_{{\rm Ideal}|x}(\overline{\Lambda})\|_1.
\end{aligned}$$

Therefore, it is enough to consider the case when the channel is $\overline{\Lambda}$ even if the used channel Λ is not a Pauli channel.

Appendix B. Proof of Lemma 1

In order to prove this lemma, we introduce several new lemmas. In the first part, i.e, Appendix B.1, we derive exact upper bounds on $P_{\rm hg}(c|k)$ given in terms of l or $s(\varepsilon)$. Then in Appendix B.2, we show that those upper bounds can also be bounded by $\varepsilon=\Phi^{-1}(s(\varepsilon))$. Finally in Appendix B.3, using the obtained results, we prove Lemma 1.

Appendix B.1. Upper Bounds on sums of $P_{\rm hg}(c|k)$

Lemma 2 *If $l\le n$ and $\frac{1}{n+l}\le\frac{k}{n+l}\le\frac{1}{2}$,*

$$\sum_{i=0}^{c}P_{\rm hg}(i|k)\le D_{n,l,k}(c), \tag{B.1}$$

where

$$\begin{aligned}
D_{n,l,k}(c):=&\sqrt{\frac{n(n+l-k)k}{(n+l)(n-k+c)(k-c)}} \\
&\times e^{\mu}2^{nh\left(\frac{k-c}{n}\right)-(n+l)h\left(\frac{k}{n+l}\right)+lh\left(\frac{c}{l}\right)},
\end{aligned} \tag{B.2}$$

$$\mu:=\frac{1}{6n}+\frac{1}{12}. \tag{B.3}$$

Proof: By using the Stirling's formula

$$n!=\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\lambda_n}\quad\text{with}\quad\frac{1}{12n+1}<\lambda_n<\frac{1}{12n}, \tag{B.4}$$

we have

$$\begin{aligned}
\frac{\binom{n}{k-c}}{\binom{n+l}{k}}=&\sqrt{\frac{n(n+l-k)k}{(n+l)(n-k+c)(k-c)}} \\
&\times e^{\mu'}\,2^{nh\left(\frac{k-c}{n}\right)-(n+l)h\left(\frac{k}{n+l}\right)}
\end{aligned} \tag{B.5}$$

where

$$\begin{aligned}\mu' &:= \lambda_n - \lambda_{n-k+c} - \lambda_{k-c} - \lambda_{n+l} + \lambda_{n+l-k} + \lambda_k \\ &< \lambda_n + \lambda_{n+l-k} + \lambda_k < \frac{1}{6n} + \frac{1}{12}\end{aligned}$$

for $\frac{1}{n+l} \leq \frac{k}{n+l} \leq \frac{1}{2}$ and $l \leq n$. Combining (B.5) with $\sum_{i=0}^{c} \binom{l}{i} \leq 2^{lh\left(\frac{c}{l}\right)}$ (see, e.g., Lemma 4.2.2 of [29]), we obtain (B.1). □

Lemma 3 *For $l \leq n$, $c \leq \bar{c}(k)$, and $\frac{k}{n+l} < \frac{1}{2}$*

$$nh\left(\frac{k-c}{n}\right) - (n+l)h\left(\frac{k}{n+l}\right) + lh\left(\frac{c}{l}\right) \leq -\frac{1}{2\ln 2}\left(\frac{c-\bar{c}(k)}{\sigma(k)}\right)^2. \quad \text{(B.6)}$$

Proof: Since $h'''(x)$ decreases monotonically, we have

$$h(x) \leq h(x_0) + h'(x_0)(x-x_0) + \frac{1}{2}h''(x_0)(x-x_0)^2 + \frac{1}{6}h'''(x_0)(x-x_0)^3. \quad \text{(B.7)}$$

(Let $\tilde{h}(x)$ be the LHS minus the RHS. It is easy to verify that $\tilde{h}(x_0) = \tilde{h}'(x_0) = \tilde{h}''(x_0) = \tilde{h}'''(x_0) = 0$ and that $\tilde{h}'''(x) = h'''(x) - h'''(x_0)$ is a decreasing function. Then by integrating $\tilde{h}'''(x)$ three times, one can show that $\tilde{h}(x) \leq 0$.) Applying inequality (B.7) for $x_0 = \frac{k}{n+l}$ and $x = \frac{k-c}{n}$, and also for $x = \frac{c}{l}$, we have

$$\begin{aligned}&nh\left(\frac{k-c}{n}\right) - (n+l)h\left(\frac{k}{n+l}\right) + lh\left(\frac{c}{l}\right) \\ &\leq \frac{1}{2}h''\left(\frac{k}{n+l}\right)\frac{n+l}{nl}(c-\bar{c}(k))^2 \\ &\quad + \frac{1}{6}h'''\left(\frac{k}{n+l}\right)\left\{\frac{1}{n^2} - \frac{1}{l^2}\right\}(\bar{c}(k)-c)^3.\end{aligned} \quad \text{(B.8)}$$

Since $h'''\left(\frac{k}{n+l}\right)$, $\bar{c}(k) - c$, and $n - l$ are all non-negative by the conditions stated in the lemma, the second term on the right hand side is non-positive. Then by noting

$$\frac{n+l}{nl}h''\left(\frac{k}{n+l}\right) = -\frac{1}{(\ln 2)\sigma(k)^2}\frac{n+l}{n+l-1} \leq -\frac{1}{(\ln 2)\sigma(k)^2},$$

we have Inequality (B.6). □

Lemma 4 *If $c \leq \bar{c}(k)$, we have*

$$\sqrt{\frac{n(n+l-k)k}{(n+l)(n-k+c)(k-c)}} \leq \sqrt{\frac{n+l}{n}} \quad \text{(B.9)}$$

Proof: Let

$$C(n,l,k,c) := \frac{n^2(n+l-k)k}{(n+l)^2(n-k+c)(k-c)}.$$

Then it suffices to show $C \leq 1$ for $0 \leq c \leq \bar{c}(k)$.

The function $f(k,c) := (n-k+c)(k-c)$ inside the square root is a concave parabola with its vertex at $c = k - \frac{n}{2}$. This means that $f(k,c) \geq \min\{f(k,\bar{c}(k)), f(k,0)\}$, and thus $C(n,l,k,c) \leq \max\{C(n,l,k,\bar{c}(k)), C(n,l,k,0)\}$. Then it is straightforward to verity $C(n,l,k,\bar{c}(k)) = 1$ and $C(n,l,k,0) \leq 1$. □

Lemma 5 *If $l \leq n$, $1 \leq k$, $c \leq \bar{c}(k)$ and $\frac{k}{n+l} \leq \frac{1}{2}$, we have*

$$\sum_{i=0}^{c} P_{\text{hg}}(i|k) \leq e^{\mu} \sqrt{\frac{n+l}{n}} \exp\left[-\frac{1}{2}\left(\frac{c-\bar{c}(k)}{\sigma(k)}\right)^2\right]. \tag{B.10}$$

Proof: Combine Lemmas 2, 3 and 4. □

Lemma 6 *If $0 \leq t$, $\bar{c}(k) - lt \leq l/2$ and $\frac{k}{n+l} \leq \frac{1}{2}$,*

$$\sum_{c=0}^{\bar{c}(k)-lt} P_{\text{hg}}(c|k) \leq \exp\left[\frac{lt^2}{2} h''\left(\frac{k}{n+l}\right)\right]. \tag{B.11}$$

Proof: According to [15],

$$\begin{aligned}\sum_{i=0}^{\bar{c}(k)-lt} P_{\text{hg}}(i|k) &\leq \left(\left(\frac{p}{p-t}\right)^{p-t}\left(\frac{1-p}{1-(p-t)}\right)^{1-(p-t)}\right)^{l} \\ &= 2^{l[h(p-t)-h(p)+th'(p)]},\end{aligned} \tag{B.12}$$

where $p = \frac{\bar{c}(k)}{l} = \frac{k}{n+l}$. Since $h''(x)$ increases monotonically for $p - t \leq x \leq p \leq 1/2$, we have

$$h(p-t) \leq h(p) + (-t)h'(p) + \frac{(-t)^2}{2} h''(p).$$

That is,

$$l[h(p-t) - h(p) + th'(p)] \leq \frac{lt^2}{2} h''(p)$$

□

Appendix B.2. Upper and Lower Bounds on $\Phi(x)$

Lemma 7 *The normal distribution function, defined in (19), is bounded as*

$$\frac{\sqrt{2}}{\sqrt{x^2+2\pi}} e^{-x^2/2} \leq \Phi(x) \leq \frac{\sqrt{2}}{x} e^{-x^2/2}. \tag{B.13}$$

Proof: According to Ref. [14], the function $\Phi(x)$ satisfies

$$\tilde{g}_{\pi}(x) e^{-x^2/2} \leq \Phi(x) \leq \tilde{g}_4(x) e^{-x^2/2}, \tag{B.14}$$

where

$$\tilde{g}_k(x) := \frac{\sqrt{2k}}{(k-1)x + \sqrt{x^2+2k}}. \tag{B.15}$$

Then it is straightforward to show that for $k, x > 0$,

$$\frac{\sqrt{2}}{\sqrt{x^2+2k}} \leq \tilde{g}_k(x) \leq \frac{\sqrt{2}}{x}. \tag{B.16}$$

Combining (B.14) and (B.16), we obtain the lemma. □

Lemma 8 *If* $\varepsilon = \Phi(s)$, *and* $2 \leq s$,

$$e^{-s^2} \leq \frac{\varepsilon}{2}. \tag{B.17}$$

Proof: From Lemma 7,

$$e^{-s^2} \leq e^{-s^2/2}\frac{\sqrt{s^2+2\pi}}{\sqrt{2}}\Phi(s) = \sqrt{\frac{(s^2+2\pi)e^{-s^2}}{2}}\,\varepsilon.$$

Then by noting $\frac{(s^2+2\pi)e^{-s^2}}{2} \leq \frac{1}{4}$ for $2 \leq s$, we obtain the lemma. □

Appendix B.3. Proof of Lemma 1

If $k/(n+l) \leq \frac{1}{2}$, by combining Lemmas 5 and 7, we obtain

$$\sum_{c=0}^{\lfloor \bar{c}-s\sigma \rfloor} P_{\rm hg}(i|k) \leq \sqrt{\frac{n+l}{n}}\sqrt{\frac{s^2+2\pi}{2}}e^{\mu}\varepsilon.$$

On the other hand, if $k/(n+l) > \frac{1}{2}$, by Lemma 6, we have

$$\begin{aligned}\sum_{c=0}^{c_{\max}} P_{\rm hg}(c|k) &\leq \sum_{c=0}^{c_{\max}} P_{\rm hg}\left(c|(n+l)/2\right) \\ &\leq \exp\left[\frac{(1/2-0.12)^2 l}{2}h''(1/2)\right] \\ &\leq e^{-\frac{2}{5}l} \leq e^{-\frac{s^2}{2}}.\end{aligned} \tag{B.18}$$

Then by using Lemma 7, we have

$$\sum_{c=0}^{c_{\max}} P_{\rm hg}(i|k) < e^{-\frac{1}{2}s^2} \leq \sqrt{\frac{s^2+2\pi}{2}}\,\varepsilon. \tag{B.19}$$

Appendix C. Proof of Theorem 1

Appendix C.1. Proof of Case 1

Since $p_{\rm sft}(k,c) = \frac{k-c}{n} \leq \frac{k}{n} \leq \frac{c_{\min}}{l} = p_{\rm smp}(c_{\min})$, we have for arbitrary $c \in [0,l]$,

$$\begin{aligned}g(k,c) &= nh\left(p_{\rm sft}(k,c)\right) - nh\left(\hat{p}_{{\rm sft},\varepsilon}\left(\max\{c+2, c_{\min}\}\right)\right) - D \\ &\leq nh\left(p_{\rm smp}(c_{\min})\right) - nh\left(\hat{p}_{{\rm sft},\varepsilon}(c_{\min})\right) - D.\end{aligned}$$

Further, from the concavity of $h(x)$ and from the monotonicity of $h'(x)$,

$$\begin{aligned}g(k,c) &\leq nh'(\hat{p}_{{\rm sft},\varepsilon}(c_{\min}))\left[p_{\rm smp}(c_{\min}) - \hat{p}_{{\rm sft},\varepsilon}(c_{\min})\right] \\ &\leq nh'(\hat{p}_{{\rm sft},\varepsilon}(c_{\max}))\left[p_{\rm smp}(c_{\min}) - \hat{p}_{{\rm sft},\varepsilon}(c_{\min})\right].\end{aligned}$$

Then by using Eq. (25) and by noting that $\left(p_{\rm smp} - \hat{p}_{\varepsilon}\right)^2 = 4\gamma\hat{p}_{\varepsilon}(1-\hat{p}_{\varepsilon})$ (see below Eq. (23)),

$$\begin{aligned}g(k,c) &\leq -(n+l)h'(\hat{p}_{{\rm sft},\varepsilon}(c_{\max}))\left[\hat{p}_{\varepsilon}(c_{\min}) - p_{\rm smp}(c_{\min})\right] - D \\ &= -(n+l)h'(\hat{p}_{{\rm sft},\varepsilon}(c_{\max}))\end{aligned}$$

$$\begin{aligned} &\times \sqrt{4\gamma}\sqrt{\hat{p}_\varepsilon(c_{\min})(1-\hat{p}_\varepsilon(c_{\min}))} - D \\ =\ & -(1+4\gamma)s(\varepsilon)\beta\sigma\left((n+l)\hat{p}_\varepsilon(c_{\min})\right) - D \\ \le\ & -\frac{\xi_{\min,\varepsilon}s(\varepsilon)^2}{\ln 2} - D. \end{aligned}$$

The last inequality follows by noting that $nc_{\min}/l \le (n+l)\hat{p}_\varepsilon(c_{\min}) \le (n+l)\hat{p}_\varepsilon(c_{\max})$, and thus $-\sigma\left((n+l)\hat{p}_\varepsilon(c_{\max})\right) \le -\sigma\left(nc_{\min}/l\right)$. Then by using Lemma 8, we have for $1 < \xi_{\min,\varepsilon}$ and $D = 1$,

$$S_{\mathrm{pa}}(k,c) \le 2^{[g(k,c)]^- + 1} \le 2e^{-\xi_{\min,\varepsilon}s(\varepsilon)^2} < \varepsilon$$

□

Appendix C.2. Proof of Case 2

This part is immediate from the following lemma.

Lemma 9 *Suppose $1 \le l \le n$, $4\gamma \le 1$. Then, for any integer k, any real number $\varepsilon > 0$ and any $c \in [\,\bar{c}(k) - s(\varepsilon)\sigma(k),\, c_{\max}\,]$, we have*

$$g(k,c) \le -\beta\left(c - (\bar{c}(k) - s(\varepsilon)\sigma(k)) + 1\right) - D, \tag{C.1}$$

with β defined in (36).

Proof: With $h(x)$ being concave, and with $\hat{p}_{\mathrm{sft},\varepsilon}(c)$ increasing monotonically,

$$\begin{aligned} g(k,c) \le\ & -nh'(\hat{p}_{\mathrm{sft},\varepsilon}(c+2))\left(\hat{p}_{\mathrm{sft},\varepsilon}(c+2) - p_{\mathrm{sft}}(k,c)\right) - D \\ \le\ & -nh'(\hat{p}_{\mathrm{sft},\varepsilon}(c_{\max}+2))\left(\hat{p}_{\mathrm{sft},\varepsilon}(c+2) - p_{\mathrm{sft}}(k,c)\right) - D. \end{aligned}$$

The quantity $\hat{p}_{\mathrm{sft},\varepsilon}(c+2) - p_{\mathrm{sft}}(k,c)$ on the right hand side can be bounded as follows. First note $\hat{p}_{\mathrm{sft},\varepsilon}(\bar{c} - s\sigma) - p_{\mathrm{sft}}(k, \bar{c} - s\sigma) = 0$ by the definition of $\hat{p}_{\mathrm{sft},\varepsilon}(c)$, given in (24) and (25). Also by the definition of $\hat{p}_{\mathrm{sft},\varepsilon}(c)$, we have that $\frac{d\hat{p}_{\mathrm{sft},\varepsilon}}{dc} \ge \frac{1}{1+4\gamma}\frac{n+l}{nl} - \frac{1}{n}$, and that $\frac{\partial p_{\mathrm{sft}}}{\partial c} = -\frac{1}{n}$ by the definition of $p_{\mathrm{sft}}(k,c)$; hence $\frac{\partial}{\partial c}(\hat{p}_{\mathrm{sft},\varepsilon} - p_{\mathrm{sft}}) \ge \frac{1}{1+4\gamma}\frac{n+l}{nl}$. Thus $\hat{p}_{\mathrm{sft},\varepsilon}(\bar{c} - s\sigma + 2) - p_{\mathrm{sft}}(k, \bar{c} - s\sigma + 2) \ge \frac{2}{1+4\gamma}\frac{n+l}{nl}$. Then for $\bar{c}(k) - s(\varepsilon)\sigma(k) \le c$, we have

$$\begin{aligned} &\hat{p}_{\mathrm{sft},\varepsilon}(c+2) - p_{\mathrm{sft}}(k,c) && \text{(C.2)} \\ &= (\hat{p}_{\mathrm{sft},\varepsilon}(c+2) - p_{\mathrm{sft}}(k,c+2)) + (p_{\mathrm{sft}}(k,c+2) - p_{\mathrm{sft}}(k,c)) && \text{(C.3)} \\ &\ge \frac{1}{1+4\gamma}\frac{n+l}{nl}\left(c - (\bar{c} - s\sigma) + 2\right) - \frac{2}{n} && \text{(C.4)} \\ &\ge \frac{1}{1+4\gamma}\frac{n+l}{nl}\left(c - (\bar{c} - s\sigma) + 1\right). && \text{(C.5)} \end{aligned}$$

□

Plugging this upper bound on $g(k,c)$ (for $D = 1$) to $S_{\mathrm{pa}}(k,c)$ (given in (13) and (14)), we obtain Case 2 of Theorem 1.

Appendix D. Proof of Theorem 3

Next we prove Theorem 3 starting from Theorem 1. In the following, $s(\varepsilon)$ is simplified to s.

Under the conditions of Case 1 of Theorem 1, inequality (34) holds independently of the normal approximation, and thus we readily see that (49) holds.

Lemma 10 *If $1 \le l \le n$, $1 \le k$, $c \le \bar{c}(k)$ and $\frac{k}{n+l} \le 1/2$, we have*

$$P_{\rm hg}(c|k) \le \frac{e^{\mu+\nu}}{\sqrt{2\pi}\sigma((n+l)c/l)} \exp\left[-\frac{1}{2}\left(\frac{c-\bar{c}(k)}{\sigma(k)}\right)^2\right], \tag{D.1}$$

with μ defined in (B.3), and

$$\nu := \frac{1}{12l} + \frac{1}{2(n+l-1)}. \tag{D.2}$$

Proof: By using the Stirling's formula (B.4), we have

$$\binom{l}{c} \le \sqrt{\frac{n}{n+l-1}} \frac{1}{\sqrt{2\pi}\sigma((n+l)c/l)} e^{\nu'} 2^{lh(c/l)}, \tag{D.3}$$

where

$$\nu' = \lambda_l - \lambda_{l-c} - \lambda_c \le \lambda_l < \frac{1}{12l}. \tag{D.4}$$

Then by combining Inequality (D.3) with (B.5) and (B.6), and by using Lemma 4, we obtain

$$P_{\rm hg}(c|k) \le \frac{e^{\mu+\frac{1}{12l}}}{\sqrt{2\pi}\sigma((n+l)c/l)} \sqrt{1+\frac{1}{n+l-1}} \exp\left[-\frac{1}{2}\left(\frac{c-\bar{c}(k)}{\sigma(k)}\right)^2\right].$$

Then by noting

$$\sqrt{1+\frac{1}{n+l-1}} \le \sqrt{\exp\left(\frac{1}{n+l-1}\right)} = \exp\left(\frac{1}{2(n+l-1)}\right),$$

we obtain the lemma. □

Lemma 11 *If $l \le n$, $1 \le c_{\min}$, $nc_{\min}/l \le k$, $\bar{c}(k) - s\sigma(k) \le c \le \bar{c}(k)$ and $\frac{k}{n+l} \le 1/2$, we have*

$$P_{\rm hg}(c|k) \le \frac{e^{\mu+\nu}}{\sqrt{1-\frac{s}{\sqrt{c_{\min}}}}} \frac{1}{\sqrt{2\pi}\sigma(k)} \exp\left[-\frac{1}{2}\left(\frac{c-\bar{c}(k)}{\sigma(k)}\right)^2\right], \tag{D.5}$$

with μ, ν defined in (B.3), (D.2)

Proof: From the definition of $\sigma(k)$, we have

$$\frac{\sigma(k)}{\sigma(k(1-s\sigma(k)/\bar{c}(k)))} \le \frac{1}{\sqrt{1-s\sigma(k)/\bar{c}(k)}}.$$

By noting that $nc_{\min}/l \leq k$, we have

$$\begin{aligned}
\frac{\sigma(k)}{\bar{c}(k)} &= \sqrt{\frac{n}{l(n+l-1)}\left(\frac{n+l}{k}-1\right)} \\
&\leq \sqrt{\frac{n}{l(n+l-1)}\left(\frac{l(n+l)}{nc_{\min}}-1\right)} \\
&= \sqrt{\frac{n}{l(n+l-1)}\left(\frac{l(n+l)-nc_{\min}}{nc_{\min}}\right)} \\
&\leq \sqrt{\frac{n}{l(n+l-1)}\left(\frac{l(n+l-1)}{nc_{\min}}\right)} \\
&\leq \frac{1}{\sqrt{c_{\min}}}.
\end{aligned}$$

Hence $1-\frac{s\sigma(k)}{\bar{c}(k)} \geq 1-\frac{s}{\sqrt{c_{\min}}}$. The assumption yields that $(n+l)c/l \geq k(1-s\sigma(k)/\bar{c}(k))$, which implies

$$\frac{\sigma(k)}{\sigma((n+l)c/l)} \leq \frac{\sigma(k)}{\sigma(k(1-s\sigma(k)/\bar{c}(k)))} \leq \frac{1}{\sqrt{1-\frac{s}{\sqrt{c_{\min}}}}}.$$

Combining this inequality with Lemma 10, we obtain Lemma 11. □

Appendix D.1. Proof of Case 2

If $\frac{k}{n+l} \geq \frac{1}{2}$, this case can be proved by exactly the same argument as in Appendix B.3 (Note here that the condition $s^2 \leq c_{\min} \leq c_{\max} \leq 0.12l$, appearing in Theorem 3, implies $\frac{5}{4}s^2 \leq l$). Hence in this subsection, we assume that $\frac{k}{n+l} < \frac{1}{2}$. We also assume that $1 \leq k$, because the case $k=0$ is already considered in Case 1 of Theorem 1.

First we divide the right hand side of (37) into three parts,

$$\begin{aligned}
&\sum_{c=0}^{c_{\max}} P_{\mathrm{hg}}(c|k)S_{\mathrm{pa}}(k,c) \\
&\leq \sum_{c=0}^{\lfloor \bar{c}(k)-s\sigma(k)\rfloor} P_{\mathrm{hg}}(c|k) + \sum_{c=\lfloor \bar{c}(k)-s\sigma(k)\rfloor+1}^{\lfloor \bar{c}(k)\rfloor-1} P_{\mathrm{hg}}(c|k)S_{\mathrm{pa}}(k,c) \\
&\quad + \sum_{c=\lfloor \bar{c}(k)\rfloor}^{c_{\max}} P_{\mathrm{hg}}(c|k)S_{\mathrm{pa}}(k,c). \qquad \text{(D.6)}
\end{aligned}$$

The first term on the right hand side can be bounded from above by Lemma 1. The second term can be bounded as

$$\begin{aligned}
&\sum_{c=\lfloor \bar{c}(k)-s\sigma(k)\rfloor+1}^{\lfloor \bar{c}(k)\rfloor-1} P_{\mathrm{hg}}(c|k)S_{\mathrm{pa}}(k,c) \\
&\leq \sum_{c=\lfloor \bar{c}(k)-s\sigma(k)\rfloor+1}^{\lfloor \bar{c}(k)\rfloor-1} P_{\mathrm{hg}}(c|k)2^{-\beta(c-(\bar{c}(k)-s\sigma(k))+1)}
\end{aligned}$$

$$\begin{aligned}
&\le \frac{e^{\mu+\nu}}{\sqrt{1-\frac{s}{\sqrt{c_{\min}}}}} \frac{1}{\sqrt{2\pi}\sigma(k)} \\
&\quad \times \sum_{c=\lfloor \bar{c}(k)-s\sigma(k)\rfloor+1}^{\lfloor \bar{c}(k)\rfloor-1} \exp\left[-\frac{1}{2}\left(\frac{c-\bar{c}(k)}{\sigma(k)}\right)^2\right] \\
&\qquad\qquad \times 2^{-\beta(c-(\bar{c}(k)-s\sigma(k))+1)} \\
&\le \frac{e^{\mu+\nu}}{\sqrt{1-\frac{s}{\sqrt{c_{\min}}}}} \frac{1}{\sqrt{2\pi}} \int_{-s}^{\infty} dx\ e^{-x^2/2} 2^{-\beta\sigma(k)(x+s)} \\
&\le \frac{e^{\mu+\nu}}{\sqrt{1-\frac{s}{\sqrt{c_{\min}}}}} I_2\left(\xi_\varepsilon(k)\right).
\end{aligned}$$

Then $I_2\left(\xi_\varepsilon(k)\right)$ appearing in the last line can be bounded by Inequality (44). (Note that the argument in the paragraph of Inequality (44) does not rely on the normal approximation.)

The third summation on the right hand side of (D.6) can be bounded as

$$\begin{aligned}
&\sum_{c=\lfloor \bar{c}(k)\rfloor}^{c_{\max}+1} P_{\rm hg}(c|k) 2^{-\beta(c-(\bar{c}(k)-s\sigma(k))+1)} \\
&\le \sum_{c=\lfloor \bar{c}(k)\rfloor}^{c_{\max}+1} P_{\rm hg}(c|k) 2^{-\beta(c-(\bar{c}(k)-s\sigma(k))+1)} \\
&\le 2^{-\beta\sigma(k)s} \le e^{-\xi_\varepsilon(k)s^2} \le \varepsilon^{\xi_\varepsilon(k)} \le \varepsilon^2.
\end{aligned}$$

Appendix E. Proof of Theorem 4:

First, we fix arbitrary $\varepsilon' > \varepsilon$. Since the function $h(x)$ and its derivative $h'(x)$ are uniformly continuous in the range $[p_{\min}, p_{\max}]$, there exists an integer N such that $\lceil nh\left(\hat{p}_{{\rm sft},\varepsilon'}(c+1)\right)\rceil + 1 \le \lceil nh\left(p_{\rm smp}(c)\right) + \sqrt{n}h'\left(p_{\rm smp}(c)\right)\sqrt{\frac{p_{\rm smp}(c)(1-p_{\rm smp}(c))(1+t)}{4t}}s(\varepsilon)\rceil$ for $n \ge N$ and $l \ge tn$. Using Theorem 1 of [30], we can choose constants C_1 and C_2 such that $P_{\rm hg}(c|k) \le \frac{1}{\sqrt{2\pi}}\int_{\zeta_c}^{\zeta_{c+1}} e^{-x/2}dx + \frac{C_1(1+\zeta_c^2)}{\sigma_{n,l}(k)}\exp(-C_2\zeta_c^2)$. Here note that the constants C_1 and C_2 are different from those defined in Theorem 1 of [30].

Using $C_3 := \int_{-\infty}^{\infty} C_1(1+x^2)\exp(-C_2x^2)$, we obtain

$$\sum_{c=0}^{np_{\max}} \frac{C_1(1+\zeta_c^2)}{\sigma_{n,l}(k)} \exp(-C_2\zeta_c^2) \min\left\{2^{-\beta(c-(\bar{c}-s(\varepsilon)\sigma))}, 1\right\} \le \frac{C_3}{\sigma_{n,l}(k)}. \tag{E.1}$$

Hence, Theorem 2 yields that

$$P_{{\rm ph},n,l} \le (1+\delta_n')\varepsilon' + \frac{C_3}{\min_{k:np_{\min}\le k\le (n+l)\left(\hat{p}_{{\rm sft},\varepsilon'}(lp_{\max}+1)\right)} \sigma_{n,l}(k)}. \tag{E.2}$$

where δ_n' is the maxumum of δ given in Theorem 2 with the condition $l \ge tn$.

Since $\min_{l:l\ge tn} \min_{k:np_{\min}\le k\le (n+l)\left(\hat{p}_{{\rm sft},\varepsilon'}(lp_{\max}+1)\right)} \sigma_{n,l}(k) \to \infty$ as $n\to\infty$, we obtain $\lim_{n\to\infty} \max_{l:l\ge tn} \frac{C_3}{\min_{k:np_{\min}\le k\le (n+l)\left(\hat{p}_{{\rm sft},\varepsilon'}(lp_{\max}+1)\right)} \sigma_{n,l}(k)} = 0$. Also we can show that $\delta_n' \to 0$.

Thus, we obtain $\lim_{n\to\infty} \max_{l:l\geq tn} P_{\mathrm{ph},n,l} \leq \varepsilon'$. Since ε' is an arbitrary real number satisfying that $\varepsilon' > \varepsilon$. Hence, $\lim_{n\to\infty} \max_{l:l\geq tn} P_{\mathrm{ph},n,l} \leq \varepsilon$. $\square$